

## 타일 기반 비균등 품질 360 영상에 관한 사용자 경험 평가 방법 연구

이순빈, 장동민, 정종범, 류일웅, Tuan Thanh Le, 류은석  
가천대학교

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

### A Study on Viewport Quality Assessment for Tile Based Mixed Quality 360 Video

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

#### 요 약

360 영상을 이용한 실감형 미디어에 대한 관심이 날로 높아지고 있지만 높은 대역폭 요구사항으로 인해 원활히 보급되지 못하고 있는 실정이다. 이에 따라 360 영상을 타일 분할 기법을 적용한 뒤 각 영역의 중요도에 따라 독립적으로 품질을 조절하여 서비스하는 타일 기반 방법이 활발히 연구되고 있다. 그러나 사용자 시점에서 타일 경계에 있는 영상 지점을 바라볼 경우 다양한 품질의 영역이 부분적으로 존재하여 사용자 경험 평가가 어려운 문제가 발생한다. 본 논문에서는 360 영상을 비균등 품질의 타일들로 구성한 후 시선 위치 데이터를 통해 사용자 시점의 영상을 렌더링하는 과정에 대해 설명한다. 또한 다양한 품질의 영역으로 존재할 수 있는 사용자 시점의 영상의 두드러지는 영역에 대해 가중치를 적용하는 방법을 제안한다. 제안한 방법을 통하여 사용자 경험을 크게 해치지 않으면서 동시에 대역폭을 낮출 수 있는 360 영상 타일의 최적 구성을 찾기 위한 방향을 제시한다.

#### 1. 서론

VR(Virtual Reality)에 대한 기대가 점점 커져감에 따라, 몰입형 미디어 시장의 규모는 폭발적으로 늘어나고 있다. 게임, 스포츠, 가상극장 등 사용자의 몰입 경험을 위한 콘텐츠와 기술 등이 점차 등장하며 향후 시장에서 지배적인 위치를 차지할 것으로 전망하고 있다. 그러나 원활한 VR 미디어 경험을 위해서는 아직 많은 현실적 문제들이 존재하며, VR 경험을 위해 사용되는 360 영상은 전 방향에 대해 영상 정보가 필요한 만큼 기존의 영상에 비해 매우 큰 해상도를 가지고 있다. 이상적인 VR 경험을 위해서는 4K (3840x1920)부터 최대 12K (11520x6480)의 해상도를 필요로 하며 [1], 현재 가능한 대역폭으로는 원활한 전송이 어려운 요구치이다.

이러한 문제를 해결하기 위하여 360 영상의 영역 중 사용자의 관심 영역(ROI, Region of Interest)만을 고화질로 인코딩하고, 나머지 부분은 저화질로 인코딩하는 사용자 시점 기반 프로세싱 방법이 활발히 연구되고 있다. 360 영상의 영역을 타일 별로 나누어 전송하는 방식은 널리 연구되어 왔으며 [2], 비교적 사용자가 경험하지 않을 부분을 선택적으로 저화질로 압축한다. 사용자는 360 영상에서 제한된 시야각(FoV, Field of View)으로 콘텐츠를 감상하므로,

상대적으로 사용자의 경험을 크게 해치지 않으면서 필요한 대역폭을 크게 낮출 수 있게 된다.

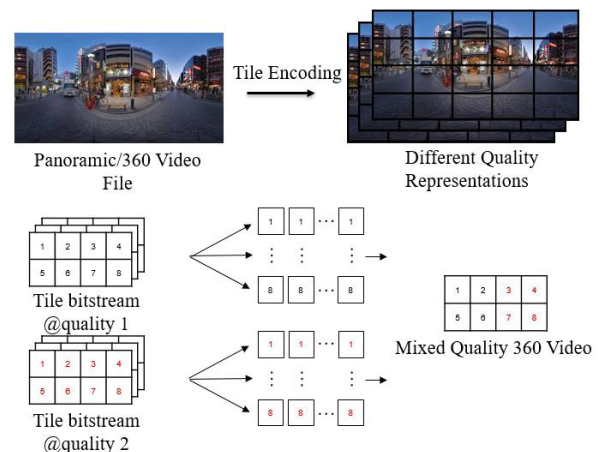


그림 1. 타일 기반 360 영상 스트리밍 개념도 [3]

위 그림은 타일 기반 360 영상 스트리밍의 개념도를

나타낸다. 서버에서는 미리 정의된 타일 개수로 360 영상을 분할한 뒤, 다양한 해상도 별로 압축하여 전송가능한 비트스트림(bitstream)을 가지게 된다. 해당 비트스트림은 360 영상 전송시에 품질별로 선택하여 전송하게 된다. 클라이언트는 받은 비트스트림을 각각의 타일로 복원하여, 하나의 360 영상으로 병합하여 사용자 시점의 영상을 감상하게 된다. 결과적으로 타일 구성 방법에 따라, 타일 영역 별로 다양한 품질의 이미지가 각각 존재하게 된다.

본 논문에서는 ERP(Equirectangular Projection) 형식의 360 영상에 움직임 참조 제한 타일(MCTS, Motion Constrained Tile Set)기법을 적용하여 타일을 분할하여 실험을 진행한다. 이후 미리 취득된 시선 위치 데이터를 이용하여 프레임 별로 유저가 실제 경험할 사용자 시점의 영상을 렌더링하고, 원본의 고화질 영상과 품질을 비교함으로써 사용자 경험에 대한 평가가 이루어진다.

위의 과정을 거쳐 구성된 타일 기반 비균등 품질 360 영상은 실제 사용자가 영상에 대해 어떻게 느끼는지에 대한 평가의 어려움이 발생하게 되며, 그에 따라 타일 영역 별로 품질을 어떻게 최적으로 구성할지에 대한 기준 또한 어려움이 존재한다. 본 논문에서는 사용자 시점의 중요한 영역에 평가 가중치를 두어 사용자 경험과 가까운 비교 방법을 제안한다.

## 2. 관련 연구

### 2.1 Tile Based Streaming with MCTS technique

보다 효율적인 360 영상의 전송을 위해, 영상을 여러 영역의 타일로 분할하여 ROI 부분을 적응적으로 스트리밍하는 기법들이 널리 연구되어 왔다 [4]. 대표적인 스트리밍 방법은 임의로 360 영상을 타일로 분할하여 이를 여러 품질로 인코딩을 거쳐 전송 시에 타일 별로 품질을 선택하게 된다.

그러나 각 타일들을 하나의 영상으로 고려하여 인코딩 시에는 타일 내의 움직임 참조 벡터가 다른 타일 영역에 존재할 경우, 각 타일들이 독립적으로 존재해야 하기 때문에 참조가 불가능하여 효율이 낮아지는 문제가 발생한다. 또한 복수개의 타일 영상을 인코딩하기 위해서는 다중 인코더와 디코더가 필요하며, 번거로운 절차와 오버헤드가 필요하게 된다.

이러한 요구사항에 따라 MPEG 표준화 단체에서는 영상의 각 타일을 공간 및 시간적 참조를 독립적으로 적용하여 인코딩하는 MCTS 기법을 제시하였다 [5],[6]. 원본 영상에서 각 타일들 간의 움직임 참조를 제한함으로써 해당 기법을 통하여 단일 인코더로부터 효율적으로 각 타일들을 독립적으로 추출하여 복원할 수 있다. 본 논문에서는 MCTS 기법을 적용한 360 영상으로 실험을 진행하였다.

### 2.2 Saliency Map Detection

컴퓨터 이미지 처리 분야에서는 사용자의 시점을 고려하였을 때 두드러지는 부분인 관심 맵(Saliency Map)에 대한 연구가 활발히 진행되어왔다. 물체의 외곽선(edge) 또는 객체 인식에 있어 중요한 정보를 나타내는 영역으로, 사용자의 시점으로 이미지를 고려할 때 부분별 압축, 사실적인 그래픽 렌더링, 이미지 특징 추출 등 방법에 크게 활용되고 있다.[7] 본 논문에서는 기존의 관심 맵 추출 방법의 결과를 이용하여, 이를 사용자 시점의 영상 평가에 이용하고자 한다. [8]

### 2.3 360 Video Dataset for Visual Attention Modeling

앞서 서술한 관심 맵 연구와 더불어 사용자 시점 기반 스트리밍이 주목받음에 따라 사용자가 어느 시점에 어떤 영역을 바라볼 것인지에 대한 예측 연구(gaze prediction) 또한 진행되고 있다 [9]. HMD(Head Mount Display) 장비의 발전으로 유저의 눈동자 시선을 기록할 수 있게 됨에 따라 이를 사전 예측 정보로 이용하거나 실시간 스트리밍에서 서버에게 정보를 전달해주는 활용 방식이 연구되고 있다.

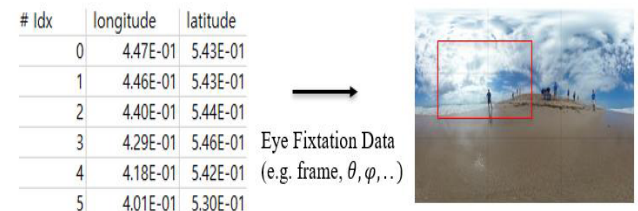


그림 2. 시선 위치 데이터 셋을 통한 사용자 시점 계산

이를 위한 시선 위치 데이터셋 [10] 또한 제안되었는데, 본 연구에서는 해당 데이터셋의 테스트 영상을 이용하여 실제 사용자 시점 영상 렌더링 과정을 구현하였다. 해당 구현을 통해 사용자가 경험할 영상을 시뮬레이션하여 사용자의 경험을 평가할 수 있다.

## 3. 관심 맵 기반 영역 평가

앞서 서술한 것처럼 MCTS 기법을 이용해 다양한 품질로 인코딩된 타일들에 대해 360 영상을 만들고, 시선 정보를 이용하여 사용자 시점 영상을 렌더링한다.



그림 3. 타일 영역 구분에 따른 비균등 품질 영역 감상

그림 3의 사용자 시점 영상에서 빨간색 영역은 낮은 품질로 사용자 시점 영상에서 눈에 띄게 사용자 경험을 저해하는 부분을 나타낸다. 이러한 해당 사용자 시점 영상을 원본과 PSNR(Peak Signal-to-noise ratio) 평가 비교를 시도해볼 수 있으며, 영역이 다소 연속적인 영역으로 품질이 달라지기 때문에 타일 영역 별로 겹침 공간( IoU, Intersection over Union) 평가 또한 고려해 볼 수 있다. 그러나 해당 방법들은 사용자 경험에 영향을 미치는 상대적인 영역을 고려하지 못한다는 한계점을 지니고 있다. 데이터 셋의 시점을 구성하여 시나리오를 따라가본 결과 사용자가 변화를 잘 눈치채지 못하는 부분과 그렇지 않은 두드러진 부분에서는 품질에 대한 평가가 달라져야 할 필요가 있다.

기존에 존재하는 ROI 기반 선택적 품질 이미지에 대한 평가 방법 역시 논의되어왔고 [11] 본 논문에서는 그 중에서도

영상의 관심 맵을 도출하여 관심 맵 영역에 가중치를 평가하는 방법을 제안한다.



그림 4. 사용자 시점 영상의 예



그림 5. 그림 4의 관심 맵(Saliency map) 도출 결과

해당 방법을 통하여 계산된 관심 맵은 객체 또는 이미지 정보가 복잡한 영역을 잘 검출해낼 수 있다. 해당 관심 맵 정보를 이용하여 두드러지는 영역의 픽셀에 대해 품질 계산 시 추가적인 가중치를 주면 사용자 시점의 경험에 중요한 부분을 상대적으로 보존하며 평가할 수 있게 된다. PSNR 평가식은 다음과 같은 수식을 통하여 계산된다.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |s(i, j) - p(i, j)|^2 \quad (1)$$

$$PSNR = \log_{10} \left( \frac{(2^{depth} - 1)^2}{MSE} \right) \quad (2)$$

품질 평가를 사용자 시점 영상에 적용하되, 관심 맵의 중요도 영역을 고려하여 특징이 두드러지는 부분에서 저화질이 일어날 경우 더 큰 벌점을 주어 평가한다. 예를 들어, 본 연구에서는 관심 맵의 값을 이진화하여 관심 맵의 픽셀 영역은 PSNR 평가 시 다음과 같이 적용한다.

$$SPSNR = \log_{10} \left( \frac{(2^{depth} - 1)^2}{(1 + t) \cdot MSE} \right) \quad (3)$$

(3)의 수식을 이용해 품질을 평가하게 되면 관심 맵의 영역에 따라 같은 픽셀 차이라도 최대 (1+t)배까지 픽셀 간의 차이값에 대해 벌점을 줄 수 있게 된다. 결과적으로 사용자가 관심 맵의 중요도가 높은 곳을 저화질로 영상을 보고 있을 경우, 중요도가 낮은 곳을 저화질로 보고 있는 경우보다 민감하게 평가할 수 있으며, 각 타일의 품질을 결정할 때 (3)의 지표를 이용하여 사용자 경험에 알맞은 품질 할당을 시도할 수 있다.



PSNR : 44.2



SPSNR(t = 10) : 38.7

기존의 비균등 품질 사용자 시점 영상에서 원본의 사용자 시점 영상과 비교했을 때에 타일 영역에 작다면 품질 평가의 값이 높게 출력되어 사용자 경험 평가에 크게 도움을 주지 못한다. 본 논문에서 제안하는 방법을 사용하여 관심 영역의 픽셀 차이에 대한 가중치를 더하게 되면 사용자 시점 영상에서 저화질의 타일에 두드러진 영역이 포함될수록 품질 평가가 낮게 측정된다.

따라서 같은 영역 넓이에서 같은 품질로 구성되어 있다 하더라도 관심 맵 정도에 따라 사용자 경험에 맞추어 품질 평가를 시행할 수 있다.

#### 4. 결론

본 논문에서는 타일 기반 비균등 품질 360 영상이 평가에 있어서 갖는 어려움을 지적하고 관심 맵을 토대로 두드러진 영역에 대하여 가중치를 두는 방법을 제안하였다. 해당 방법을 통하여 360 영상에서 각 타일의 품질을 최적으로 결정하는데 도움을 줄 수 있다. 사용자 경험 평가에 있어 제안하는 방법 이외에도 개선될 여지가 있으며, 어떤 평가 방법이 사용자 경험을 더 반영할 수 있을지 추후 실제 사용자의 의견(MOS, Mean Opinion Score)을 통해 검증할 계획이다.

#### Acknowledgement

이 논문은 2018년도 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] Hosseini, Mohammad and Viswanathan Swaminathan. "Adaptive 360 VR Video Streaming: Divide and Conquer." 2016 IEEE International Symposium on Multimedia (ISM) (2016): 107-110.
- [3] Miska Hannuksela. 2017. OMAF: viewport dependent video coding schemes. 117th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG 117/m39864.
- [4] Petrangeli, Stefano, Viswanathan Swaminathan, Mohammad Hosseini and Filip De Turck. "An HTTP/2-Based Adaptive Streaming Framework for 360° Virtual Reality Videos." ACM Multimedia (2017).
- [5] Jangwoo Son, Dongmin Jang, and Eun-Seok Ryu. "Implementing 360 video tiled streaming system."

Proceedings of the 9th ACM Multimedia Systems Conference. ACM, 2018. p. 512–524.

- [6] Jangwoo Son, Dongmin Jang, and Eun–Seok Ryu. "Implementing motion–constrained tile and viewport extraction for VR streaming." Proceedings of the 28th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video. ACM, 2018. P. 61–66.
- [7] Zünd, Fabio, Yael Pritch, Alexander Sorkine–Hornung, Stefan Mangold and Thomas R. Gross. "Content–aware compression using saliency–driven image retargeting." 2013 IEEE International Conference on Image Processing (2013): 1845–1849.
- [8] Itti, Laurent, Christof Koch and Ernst Niebur. "A Model of Saliency–Based Visual Attention for Rapid Scene Analysis." IEEE Trans. Pattern Anal. Mach. Intell. 20 (2009): 1254–1259.
- [9] Kümmerer, Matthias, Thomas S. A. Wallis and Matthias Bethge. "DeepGaze II: Reading fixations from deep features trained on object recognition." CoRR abs/1610.01563 (2016): n. pag.
- [10] Gutiérrez–Cillán, Jesús, Erwan J. David, Antoine Coutrot, Matthieu Perreira Da Silva and Patrick Le Callet. "Introducing UN Salient360! Benchmark: A platform for evaluating visual attention models for 360° contents." 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX) (2018): 1–3.
- [11] Rai, Yashas, Patrick Le Callet and Philippe Guillotel. "Which saliency weighting for omni directional image quality assessment?" 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX) (2017): 1–6.