

특집논문 (Special Paper)

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

<https://doi.org/10.5909/JBE.2019.24.6.1053>

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

사용자 시선 예측을 통한 360 영상 타일 기반 스트리밍 시스템

이 순 빈^{a)}, 장 동 민^{b)}, 정 중 범^{b)}, 이 상 순^{a)}, 류 은 석^{b)†}

Tile-Based 360 Degree Video Streaming System with User's gaze Prediction

Soonbin Lee^{a)}, Dongmin Jang^{b)}, Jong-Beom Jeong^{b)}, Sangsoon Lee^{a)}, and Eun-Seok Ryu^{b)†}

요 약

최근 360 영상에 대한 관심이 증대됨에 따라, 이러한 360 영상을 보다 효율적으로 전송하기 위해 하나의 360 영상을 여러 개의 타일로 나누어 전송하는 타일 기반 스트리밍이 활발히 연구되고 있다. 본 논문에서는 타일 기반 스트리밍 시나리오에서 사용자 시점에 대응하는 고화질 360 영상 전송을 위해, 기존 네트워크 모델로 생성된 중요도 맵(Saliency map)을 타일 기반 스트리밍에 적용하여 각 위치의 타일의 품질을 할당하는 시스템을 제안한다. 각 타일들을 독립적으로 부호화하기 위해 motion constrained tile set (MCTS) 기법을 적용함과 동시에 Salient360! 데이터셋으로 사용자 시선 시나리오를 토대로 사용자 시점 영상으로 복원하여 검증한 결과, 제안된 시스템을 기반으로 360 비디오 영상을 전송하면 기존 high-efficiency video coding (HEVC)을 사용하여 전송했을 때보다 사용자 시점의 영상은 큰 손실 없이 최대 23%의 BD-rate 효율을 보임을 확인하였다.

Abstract

Recently, tile-based streaming that transmits one 360 video in several tiles, is actively being studied in order to transmit these 360 video more efficiently. In this paper, for the transmission of high-definition 360 video corresponding to user's viewport in tile-based streaming scenarios, a system of assigning the quality of tiles at each tile by applying the saliency map generated by existing network models is proposed. As a result of usage of Motion-Constrained Tile Set (MCTS) technique to encode each tile independently, the user's viewport was rendered and tested based on Salient360! dataset, streaming 360 video based on the proposed system results in gain to 23% of the user's viewport compared to using the existing high-efficiency video coding (HEVC).

Keyword : 360 Video, HEVC, Saliency Map, Motion-Constrained Tile Set(MCTS)

a) 가천대학교 컴퓨터공학과(Gachon University, Department of Computer Engineering)

b) 성균관대학교 컴퓨터교육과(Sungkyunkwan University (SKKU), Department of Computer Education)

† Corresponding Author : 류은석(Eun-Seok Ryu)

E-mail: esryu@skku.edu

Tel: +82-2-760-0677

ORCID: <https://orcid.org/0000-0003-4894-6105>

※ 이 논문의 연구 결과 중 일부는 "ACM NOSSDAV2019"에서 발표한 바 있음.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2019-2017-0-01630).

※ This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2017-0-01630) supervised by the IITP(Institute for Information & communications Technology Promotion)

· Manuscript received October 7, 2019; Revised November 4, 2019; Accepted November 4, 2019.

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

1. 서론

실감형 미디어, 즉 virtual reality (VR)에 대한 요구가 커짐에 따라, 모바일 또는 여러 단말에서 실감형 미디어를 원활히 감상하려는 시도들이 늘어나고 있다. 특히 보다 몰입감 있는 콘텐츠를 위해 전방위로 영상을 취득가능한 360 카메라들이 보급되고 있으며, 이를 감상하기 위한 head-mounted display (HMD) 기기 또한 꾸준히 발전하고 있다. 그러나 아직까지 현재 360 비디오 스트리밍 기술로는 일반적인 시나리오에서 사용자가 충분한 정도의 몰입형 경험을 제공하기 어려우며, 이에 따라 대역폭과 연산능력을 효율적으로 조절할 수 있는 360 비디오 스트리밍 기술에 대한 요구가 커지고 있다. 특히 원활한 수준의 감상을 위해서는 최대 약 12K(11520x6480)까지 비디오 해상도가 요구된다고 보고된 바 있다^[1]. 이러한 수준의 대역폭을 감당하기 위해, 효율적인 360 비디오 스트리밍을 위한 비대칭 코어 프로세싱 기반 타일 분할 및 할당 시스템^{[2][3]}, 모바일 장치를 위한 mmWave를 이용한 데이터 오프로딩^{[4][5]} 방법 등이 제안되었다.

본 논문에서는 사용자가 360 영상을 감상할 때는 임의의

순간에 항상 제한된 시야각(Field of View, FoV)으로만 콘텐츠를 감상할 수 있다는 점에 착안하여 연구가 진행중인 사용자 시점(Viewport) 기반 스트리밍 기법에 대한 내용을 다룬다. 사용자 시점 기반 스트리밍은 360 영상의 영역 중 사용자의 관심 영역(Region of Interest, RoI) 또는 사용자가 현재 바라보고 있는 지점만을 고화질로 전송하고, 상대적으로 중요하지 않은 나머지 부분은 저화질로 전송하는 방법을 통해 사용자의 경험을 크게 해치지 않으면서 전체 영상의 대역폭을 줄이는 방법이다.

일반적으로 사용자 시점 기반 스트리밍에서 HEVC 병렬 부호화 단위인 타일을 기준으로 전체 영상을 나누어 여러 품질 단계로 부호화하고, 전송 시에는 사용자의 시점에 맞도록 각 품질의 타일 비트스트림(bitstream)들을 요청 후 복호화하여 결과적으로 영상의 품질이 비균등한 360 영상을 얻게 된다. 따라서 사용자의 관심 영역을 정확하게 예측하여 전송할 수 있다면, 전송 대역폭을 낮추면서도 사용자의 경험을 크게 향상시킬 수 있다. 그림 1은 기본적인 타일 기반 360 영상 스트리밍에 대한 개념도를 보여준다.

또한 기존에 존재하는 중요도 맵(Saliency map)을 추론하는 딥러닝 네트워크 모델을 타일 기반 360 영상 스트리밍

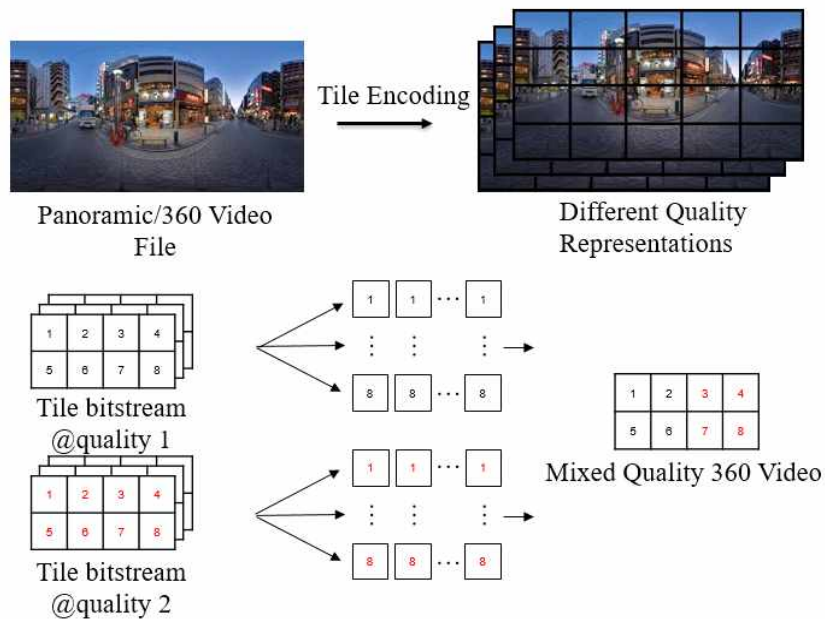


그림 1. 타일 기반 360 영상 스트리밍 개념도

Fig. 1. Conceptual diagram of the tile-based 360 video streaming system

시나리오에 적용하여, 일반적인 상황에서 보다 사용자의 시선에 맞도록 각 타일에 할당될 품질을 계산하여 개선된 360 영상 스트리밍에 대한 효율 검증을 진행하였다. Salient360! 데이터셋을 사용하여 직접 시선 위치 정보를 따라 사용자 시점을 렌더링하여 품질을 비교함으로써, 단순히 전체 영상에 대해서가 아닌 실제 사용자 경험에 대한 평가가 이루어진다.

본 논문은 시선 예측을 활용한 타일 기반 360 비디오 영상 전송을 위한 시스템을 제안한다. 타일 기반 압축 방법에 있어서 각 타일의 독립성을 보장하는 모션 참조 제한 기법(MCTS)를 적용하고, 예측된 중요도 맵을 통하여 중요도가 높은 위치의 타일에 대해 더 높은 품질을 할당하는 시스템에 대해 설명한다.

본 논문의 구성은 다음과 같다. 2 절에서는 관련 연구로 타일 기반 압축에서의 모션 참조 제한 기법(MCTS)과 중요도 맵 예측 방법에 대하여 설명한다. 3 절에서는 예측된 중요도 맵을 토대로 타일들의 품질을 할당하는 전체적인 알고리즘과 시스템에 대하여 서술한다. 4 절에서는 제안하는 시스템의 평가방법과 결과에 대해 설명하고 마지막으로 5 절에서는 본 논문의 결론을 서술한다.

II. 관련 연구

1. MCTS(Motion Constrained Tile Set)

사용자가 바라보고 있는 영역인 뷰포트에 대응하며 효율적인 전송을 위해 전체 영상을 임의의 영역으로 분할하여 부호화한 뒤, 부가 정보를 전송하는 방법인 사용자 시점 적응적 스트리밍(Viewport Adaptive Streaming)에 대해서는 여러 연구가 진행되어온 바 있다^[6,7]. 그러나 이러한 연구의 대표적인 방법은 전체 시스템 단계에서 두 가지 문제가 발생한다^[8]. 분할된 영상들을 압축된 비트스트림 형태로 전송하기 위해서는 각자 하나의 입력 영상이 되어 부호화되기 때문에 다중 인코더를 요구하며 마찬가지로 부호화된 비트스트림을 복호화하기 위해 다중 디코더가 필요한 오버헤드가 발생하게 된다. 또한 기존 HEVC의 타일 부호화 과정에서는 하나의 타일이 독립적으로 부호화되어 전송될 때, 전

송되지 않은 인접한 타일 영역에 대한 참조 문제가 발생한다.

따라서 국제 비디오 압축 양대 표준화 기구인 MPEG (ISO/IEC JTC1/SC29/WG11)과 VCEG(ITU-T Question6/WP3/SG16)으로 구성된 Joint Collaborative Team on Video Coding (JCT-VC)에서는 부호화기 차원에서 각 타일의 비트스트림들이 독립적으로 처리될 수 있도록 타일 간의 시간적 움직임 정보 참조를 제한하는 MCTS 기법을 논의한 바 있다^[9]. 이러한 MCTS 기법이 적용되어 부호화된 비트스트림은 타일의 독립성이 보장되어 원하는 영역의 타일들을 추출(Extract)할 수 있게 되며, MCTS 기법으로 부호화한 비트스트림을 미리 정의된 Supplemental Enhancement Information (SEI) 메시지를 참조하여 사용자 시점의 영상 렌더링을 위한 렌더러가 제안되기도 하였다^[10]. 이러한 MCTS 타일 정보를 포함하는 Extraction Information Sets (EIS) SEI와 함께 단일 복호화기로 다양한 영역에서 구성한 타일의 비트스트림을 추출하고 생성해낼 수 있으며, 레퍼런스 소프트웨어인 HM 16.18버전부터 MCTS 기법과 추출기(Extractor)에 대한 구현체가 내장되어 사용 가능하다^[11].

MCTS 기법은 첫번째로 시간적 참조인 Temporal inter prediction (TIP) 과정에서 eight-tap 필터 영역 내 보간이 진행될 때, 다른 타일 영역의 픽셀을 사용하게 되는 문제로 인해 이러한 보간 영역을 참조 범위에서 제외한다. 두번째로는 화면 간 예측 참조에서 모션 정보 오차를 줄이기 위해 사용하는 개선된 모션 벡터 예측 방식인 Advanced Motion Vector Prediction (AMVP)와 Merge 모드에서 참조 후보(Candidate)가 현재 타일 바깥에 존재할 수 있는 문제로 인해 타일 간 수직 경계의 참조 후보를 제외한다. 그림 2는 모션 참조 정보 예측 과정에서 타일 간 수직 경계에서의 참조 후보에 대한 문제를 나타낸다.

따라서 MCTS 기법이 적용되어 부호화된 타일들은 각 타일들이 참조할 시간적 움직임 정보가 제한되기 때문에 일반적인 HEVC 대비 부호화 효율에 있어 3~6%의 손실이 나타나는 결과가 보고된 바 있다^[12]. 그러나 MCTS 기법을 적용하면 사용자 시점 기반 스트리밍에서 보다 적은 수의 복호화기로 영상을 복원할 수 있으며, 원본 영상 비트스트림에서 다양한 영역을 유연하고 빠르게 추출할 수 있다는

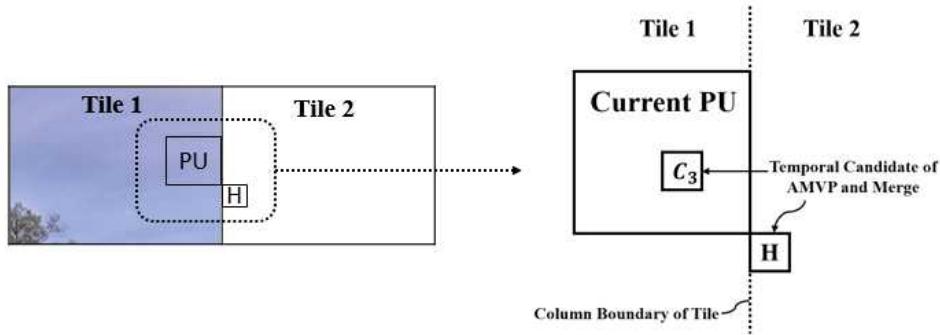


그림 2. AMVP와 Merge 모드에서의 타일 바깥의 참조 후보 문제
 Fig. 2. Reference candidate issue about the outside tiles in AMVP and Merge mode

장점을 지닌다. MCTS 기법을 이용해 나머지 영역을 전송하지 않고 사용자 시점만을 전송하면 최대 50% 정도까지 비트레이트를 절약할 수 있으며^[10], 나머지 시점을 다운샘플링하여 저화질로 보낸 일반적인 상황에서도 약 30%의 절약율을 얻는다는 것이 확인되었다^[12].

2. Saliency prediction

이러한 사용자 시점 기반 스트리밍 연구와 함께 360 영상 감상 시나리오에서 사용자의 시선을 예측하거나, 또는 그 정보를 활용하려는 연구 또한 진행되고 있다^[13]. 사용자의 관심 영역(ROI)을 예측하는 중요도 맵(Saliency map)을 검출하는 것은 전통적인 컴퓨터 비전 문제로, 기본적으로 픽셀 간의 관계를 분석하거나 외곽선(edge) 검출 등 다양한 특징들을 사용하여 이미지 또는 영상에서 사람이 인지하는 두드러지는 부분을 찾기 위한 연구가 선행되었다^[14]. 이러한 중요도 맵 검출 문제는 딥러닝 기반 방법으로 인해 현재 발전을 거듭하며 상당한 수준의 결과를 보여주고 있으며,

객체 인식(Object Recognition) 또는 분할(Segmentation) 등의 컴퓨터 비전 문제에도 활용되고 있다^[15].

360 영상 분야에서도 사용자의 눈동자와 시선을 추적하고 기록할 수 있는 eye-tracking HMD 장비들이 도입됨에 따라 데이터 주도적(Data-driven) 중요도 맵 예측에 대한 방법들이 연구되고 있다. 즉, 이미지나 영상에서 픽셀이나 내재적인 특징들만을 고려하여 검출하는 방법과 더불어 실제 사용자의 시선 기록 데이터까지 학습하여 이후 새로운 영상에서 어느 영역에 사용자 시선이 머무는지에 대한 예측을 진행한다^[16]. 대표적으로 Salient360!이라는 플랫폼은 360 영상에 대한 중요도 맵 예측 네트워크 모델들을 평가하고 비교하기 위해 360 영상과 데이터 셋을 제공하며 여러 모델들이 제안된 바 있다^[17]. 본 논문에서도 Salient360! 데이터 셋을 사용하여 전체 시스템 평가를 진행하였다.

그림 3은 각각 좌측은 원본 이미지, 중앙은 실제 기록된 사용자의 시선 맵, 우측은 360 aware saliency 모델로 예측한 중요도 맵 결과를 보여준다^[18]. 해당 모델은 기존에 존재하는 saliency 예측 모델 3가지^[19,20,21]의 결과를 종합한 뒤

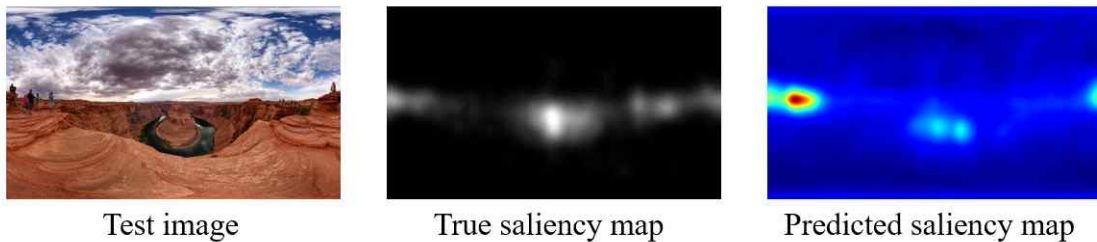


그림 3. 중요도 맵 예측의 예시 : 원본 이미지(좌), 실제 시선 맵(중앙), 예측된 중요도 맵(우)
 Fig. 3. Example of saliency map prediction : Test image(left), True saliency map(center), Predicted saliency map(right)

360 영상에 대한 왜곡 처리, 중심부에 대한 가중치 보정을 적용하여 360 영상에서 saliency 예측이 자연스럽게 이루어 지도록 고려한 모델이다. 본 논문에서는 해당 모델을 사용해 예측한 중요도 맵을 스트리밍에 사용하여 품질을 할당 하여 진행한다.

III. 사용자 시선 예측 기반 타일 스트리밍 시스템

본 논문은 위 관련연구에서 서술한 2가지 측면을 고려한 사용자 시선 예측 기반 타일 스트리밍 시스템을 제안한다. 제안하는 시스템의 주요 모듈은 2가지로 나뉘어진다. 3.1 절에서는 예측된 중요도 맵을 토대로 타일 품질을 할당하는 알고리즘에 대해 설명하고, 3.2 절에서는 예측된 중요도 맵이 일반적인 시나리오에 적합하지 않음을 지적하고 중심부에 보정을 더한 알고리즘에 대해 설명한다.

1. 타일 품질 할당 알고리즘

그림 4는 전체적인 타일 스트리밍 시스템의 구조를 나타낸다. 360 영상에 대해 사용자 시점 기반 전송 시스템을 위

하여, MCTS 기법으로 미리 부호화된 여러 품질의 타일들의 비트스트림을 가지고 있다. 또한 360 영상에 대해 네트워크 모델로 중요도 맵을 예측하는 작업을 거쳐 영상에 해당하는 중요도 맵을 갖고 있게 된다. 이후 대역폭 등 여러 상황에 따라 중요도 맵을 토대로 알고리즘을 통하여 타일 별로 품질을 선택해 부호화하여 결과적으로 하나의 여러 품질의 타일들로 구성된 하나의 360 영상을 만들게 된다. 따라서 해당 360 영상은 중요도 맵 예측 결과에 따라 사용자가 자주 보게 될 영역의 타일은 상대적으로 높은 품질이면서도, 그렇지 않은 부분은 낮은 품질로 대역폭에 유연하게 대처할 수 있다.

타일 품질 할당 알고리즘은 기본적으로 단순히 중요도 맵 영역의 값을 따라 우선순위를 주며, 이 우선순위로 진행하며 타일들의 품질을 대역폭을 넘지 않는 선에서 지속적으로 품질 단계를 올리는 방식으로 이루어진다. 예를 들어, 타일의 품질 정도에 해당하는 양자화 계수(Quantization Parameter, QP)에 따라 4단계로 이루어져 있다고 할 때 먼저 임의의 대역폭 Bandwidth에 대하여, 가장 낮은 단계의 품질로 모든 타일들을 할당하여 Bandwidth와 비교한다. 만약 Bandwidth가 더 남아있을 경우, 타일들을 중요도 맵을 토대로 정렬하여 큐에 넣은 뒤 Bandwidth를 초과하지 않는 선에서 타일의 품질을 다음 단계로 높이는 작업을 반복한

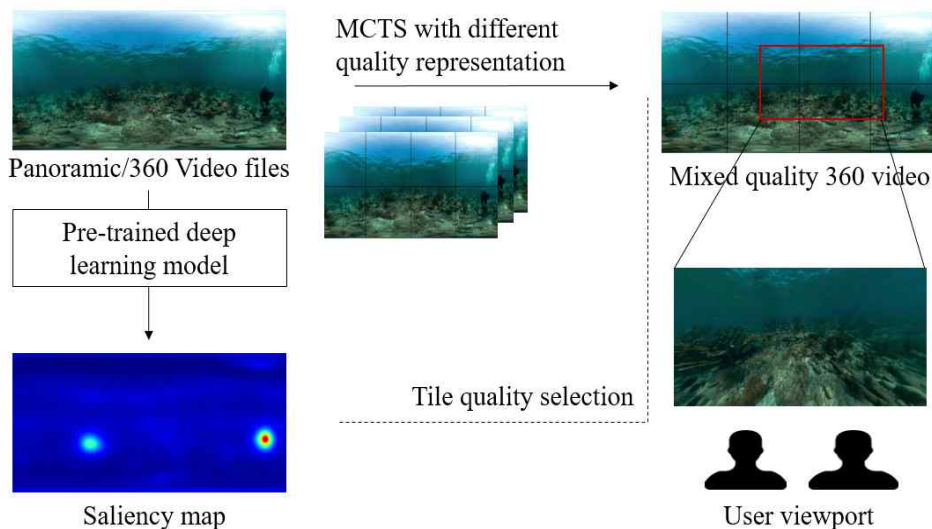


그림 4. 제안하는 사용자 시점 기반 스트리밍 시스템의 개념도
 Fig. 4. Diagram of the proposed viewport-based streaming system

다^[22].

HMD 장비를 통하여 실시간으로 사용자 시선의 정보를 받아 품질을 선택하는 서버-클라이언트 기반 스트리밍 또한 고려해볼 수 있지만, 본 논문에서는 타일 기반 스트리밍에서의 중요도 맵 활용에 대해 초점을 맞춘다. 임의 접근(Random Access, RA)이 가능한 사용자 시점 기반 스트리밍은 영상 복호화 시에 부호화되었던 이전 참조 정보의 연관성, 즉 group of pictures (GOP) 포함 문제로 사용자 시점 변경 지연(Viewport Switching Delay) 문제가 존재한다. 더불어서 eye-tracking 센서가 달려있지 않은 일반적인 HMD 장비 계측의 한계, 네트워크 지연 등의 문제들 또한 존재한다. 따라서 사전적으로 사용자 관심 영역을 찾아내어 대응하는 것 또한 관건이며, 점차 이러한 문제들을 해결하기 위해 사용자 시점 변경 지연을 해결하기 위한 연구가 이루어지고 있다^[23].

2. 360 영상 특성에 따른 개선된 우선순위 알고리즘

그러나 위와 같은 타일 품질 할당 알고리즘을 단순히 적용하기에는 문제가 존재한다. equirectangular projection (ERP)영상의 경우 그 특성상 위 아래에 왜곡이 발생하며, 전체 360 영상에서 상대적으로 작은 부분을 차지하게 된다. 본 논문에서 적용한 360 aware saliency 모델은 이미 가운데 부분에 대한 가중치가 어느정도 고려되어 있으나, 모델 자체의 보정만으로는 충분치 않다. 또한 테스트 영상에 따라 컴퓨터 그래픽(computer graphics, CG)이 포함되거나, 위 아래 부분에 중요도 맵 예측이 지나치게 높게 나오는

문제가 있다. 해당 모델은 360 이미지나 영상이 아닌 2D 콘텐츠에 대해 학습을 진행하여 360 영상에 대해 보정하는 방식으로 시행되기 때문에, ERP 영상에서 왜곡되어 나타나는 부분에 대해 경우에 따라 잘 예측하지 못하는 문제가 여전히 남아있는 것으로 판단된다. 그림 5는 CG를 포함하는 Salient360! 테스트 영상에 대한 결과를 나타낸다. 실제 영상의 내용이 진행되는 영역은 중심부 영역이므로, 일반적인 스트리밍 시나리오에서 적합하지 않다.

따라서 가운데 영역에 추가적인 우선순위를 주는 알고리즘을 제안한다. 전체 Bandwidth중에서 80% 를 가운데 영역만을 먼저 고려하여 기존의 알고리즘을 진행하고, 이후에 전체 영역에 대해 기존의 알고리즘을 진행하는 방식으로 품질 할당이 이루어진다. 따라서 반드시 중심부 영역이 80% 이상의 Bandwidth를 먼저 차지하도록 품질 결정이 이루어지게 된다. 테스트 영상에 대한 실험을 통해 수치를 결정하였으며, 선행 연구 결과 또한 해당 수치를 뒷받침한다^[24].

IV. 구현 내용 평가 및 결과

본 논문에서는 앞서 언급한 Salient360! 데이터 셋의 영상과 영상들에 해당하는 사용자 시선 기록 정보를 사용하여 시스템을 평가한다. 인코딩과 디코딩에는 HEVC test model (HM) 16.20 버전을 사용하였으며^[25], MCTS 기법을 적용하기 위해 타일 단위를 슬라이스와 1:1로 매칭시켜 인코딩을 진행한다. 타일 경계 간(cross-tile border)에 의존성

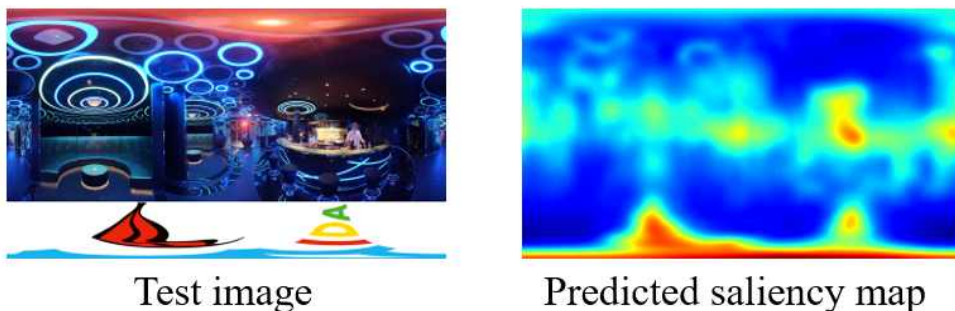


그림 5. 적합하지 않은 중요도 맵 예측의 예
Fig. 5. Example of an unsuitable saliency map prediction

표 1. 360 비디오 테스트 시퀀스 정보
 Table 1. Information of 360 video test sequences

Name	Resolution	Bitstream size (kbps)	Frame length	Frame rate	Category
PortoRiverside	3840×1920	6480	128	30 fps	low-motion
Ocean	3840×1920	7550	128	30 fps	low-motion
Turtle	3840×1920	7758	128	30 fps	low-motion
Waterpark	3840×1920	15438	128	30 fps	high-motion
UnderwaterPark	3840×1920	15336	128	30 fps	high-motion
Bar	3840×1920	22435	128	30 fps	high-motion

이 형성될 수 있는 루프내 필터링(In-loop filtering) 기능을 제외하였다. 루프내 필터링은 인코더, 디코더 내부에서 동작하는 필터링으로 추후 MCTS 기법의 경우, 품질 개선을 위해 디코더 외부에서 동작하는 필터링을 포함할 수 있다. 품질의 단계는 4단계로 각 QP 값 22, 27, 32, 37로 구성되었다. 타일의 구성 또한 4가지(4×4, 6×3, 8×4, 12×8)로 구성되며 움직임이 많은 High-motion 카테고리, 상대적으로 움직임이 덜한 Low-motion 카테고리로 영상을 나누어 평가를 진행한다. 대략적인 지표로 일부 프레임을 인코딩했을 때의 비트스트림 사이즈를 함께 나타내었다. 표 1은 구체적인 테스트 영상들의 정보를 나타낸다. 이 중 Bar 영상은 그림 5와 같이 화면 내에 CG를 포함한다.

각기 다른 품질이 할당된 타일들로 구성된 비균등 360 영상은 아직까지 객관적인 평가 지표가 없는 상황이며, 360 영상 전체에 대한 최대 신호 대 잡음비(Peak Signal-to-noise

ratio, PSNR)를 계산한다면 단순히 저품질 타일의 개수에 따라 계산되어 사용자가 직접 감상할 부분의 사용자 시점 시나리오를 반영하지 못한다. 따라서 본 연구에서는 Salient 360!에 있는 사용자 시선 기록 데이터들을 이용하여 해당 시선 기록의 위치 정보를 토대로 사용자가 직접 보게 될 영상을 렌더링한다. 그리고 렌더링된 사용자 시점을 원본의 시점과 PSNR을 비교하여 실제 사용자가 경험할 화면에 대해 손실이 얼마나 일어나는지에 대해 평가한다. 그림 6은 시스템 평가방법까지 포함한 전체 시스템의 구조도를 나타낸다. 이 때 타일 품질을 선택하는 알고리즘은 여러 알고리즘들로 대체될 수 있다.

제안하는 시스템에서는 인코딩 GOP단위인 16프레임마다 중요도 맵을 토대로 타일들의 품질을 재계산하여 다시 선택한다. 또한 렌더링하는 사용자 시점의 크기는 90×90이며, 시선 데이터셋은 해당하는 프레임 번호, 360 영상에

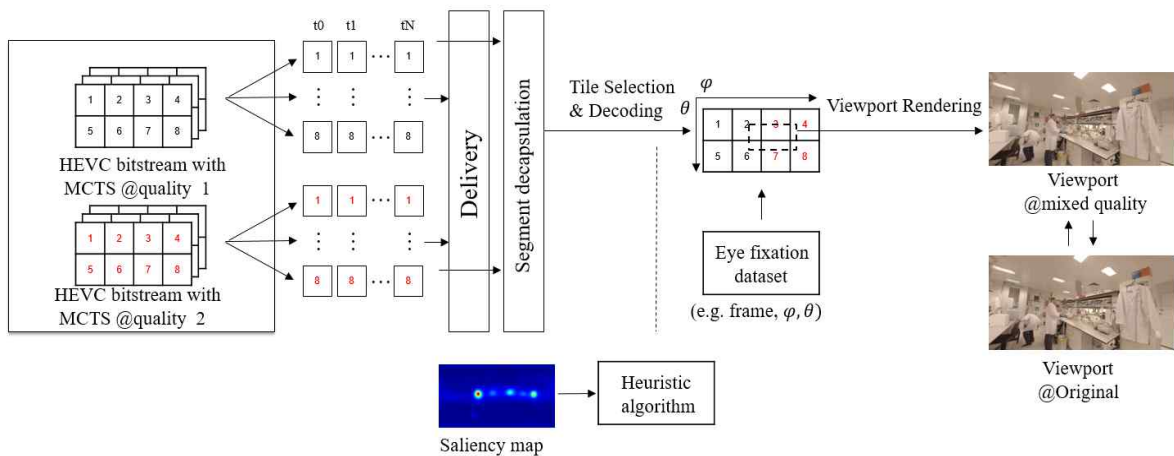


그림 6. 전체적인 스트리밍 시스템의 구조도
 Fig. 6. System overview of MCTS streaming using saliency map

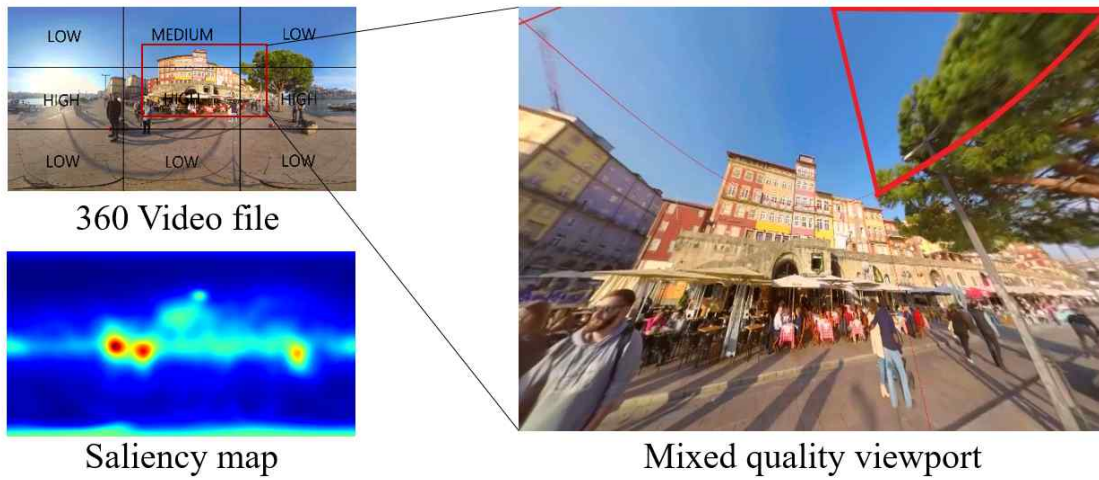


그림 7. 렌더링된 비균등 품질의 사용자 시점 이미지 예시
 Fig. 7. Example image of the Rendered mixed quality viewport

서의 위치에 해당하는 경도와 위도 값이 기록되어 있다. 영상에 따라 10~50명까지의 시선 데이터가 기록이 되어 있으며, 평가에서는 해당 기록들에 대해 평균 수치를 계산하기로 한다. 그림 7은 사용자 시점 렌더링의 한 예시를 나타내며, 붉은 테두리로 표시된 영역은 낮은 품질로 선택되어 사용자 시점에서 두드러져 보이는 타일 영역이다.

표 2에서 나타나듯이, 중심부 영역에 비트레이트 가중치를 부여한 알고리즘은 모든 경우에 대해 1~3%까지의 BD-rate 개선이 있었다. 일반적으로 타일의 개수가 많아질수록 그에 따른 오버헤드로 인한 비트레이트 증가 현상이 있지만, 예측 맵을 토대로 좀 더 세밀하게 사용자 시점의 타일 품질을 선택할 수 있게 된 만큼 상쇄하는 것으로 보인다. 그림 8에서는 표 2의 수치를 토대로 RD-curve를 나타낸

다.

위 수치는 사용자 각 개인의 경우마다 시선에 대한 확실한 예측을 진행할 수는 없기에 반드시 위와 같은 성능 향상을 보일 것이라는 보장은 없다. 따라서 이 부분에 대해서는 더 많은 사용자들의 표본과 중요도 맵의 예측 결과 판단 등 더 많은 연구가 필요하다. 그러나 평균적인 사용자의 시나리오에서 예측할 수 있는 성능 향상 지표로 사용될 수 있으며, 본 연구와 같이 향후 스트리밍이나 방송 환경에서 객체 인식, 중요도 맵 예측 등 미리 도출된 영상 정보들을 토대로 메타데이터를 함께 전송하여 타일들을 효율적으로 선택하여 추출할 수 있도록 선택 과정을 도울 수 있다.

표 2. 각 MCTS 타일 별 중요도 맵 예측 품질 할당 알고리즘의 비트레이트 비교 표
 Table 2. Bitrate comparison for MCTS technique with each algorithm based on saliency map

Method	Category	Tile Scheme / BD-rate (%)			
		4×4	6×3	8×4	12×8
Algorithm 1	High-motion	18.3	19.1	20.5	22.1
	Low-motion	16.2	16.8	17.4	18.0
	Average	17.2	17.9	18.9	20.0
Algorithm 2 (Center adaptation)	High-motion	20.1	19.7	22.4	26.3
	Low-motion	16.2	17.5	18.2	20.1
	Average	18.1	18.6	20.3	23.2

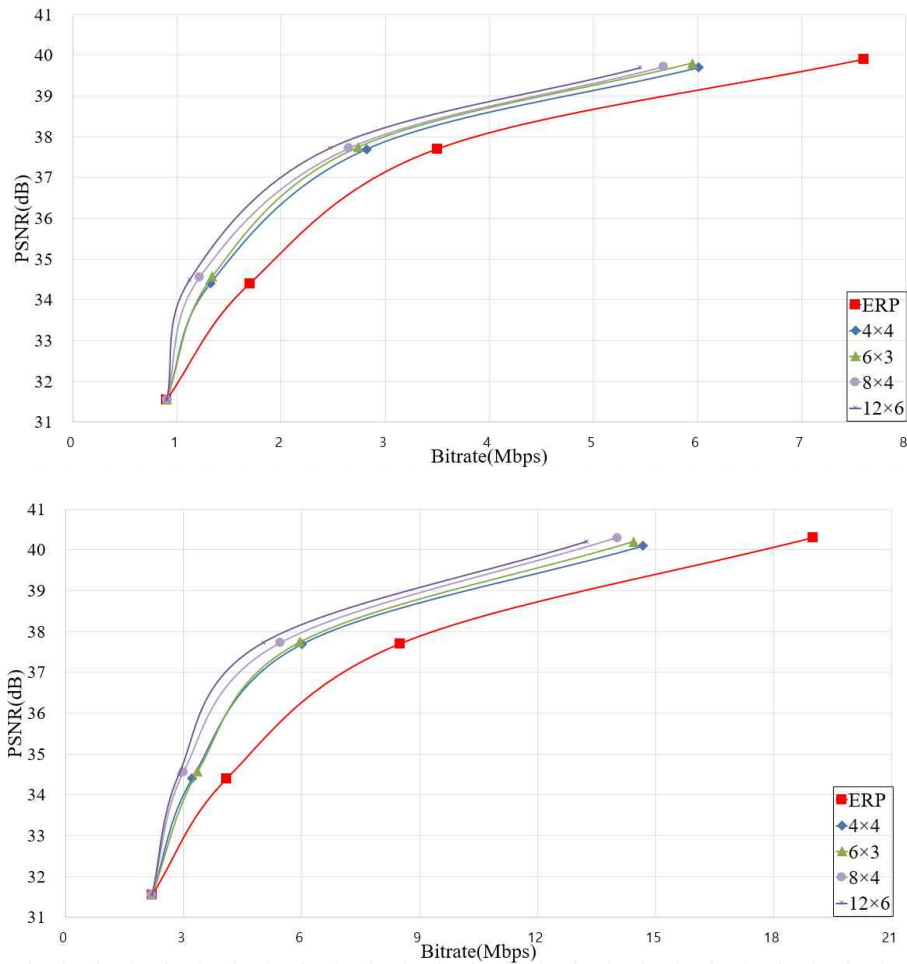


그림 8. 제안하는 시스템의 rate-distortion(RD) curve : 정적 영상 범주(위), 동적 영상 범주 (아래)
 Fig. 8. Rate-distortion(RD) curve of the proposed system : Low-motion Sequence Class(top), High-motion Sequence Class (bottom)

V. 결론

본 논문은 사용자 시선 예측을 통한 360 영상 타일 기반 스트리밍 시스템을 제안한다. 중요도 맵을 예측하는 네트워크를 활용하여 360 영상 감상 시나리오에서 타일 별로 유연하게 품질을 선택할 수 있으며, 타일 간의 독립성을 보장하는 MCTS 기법이 적용되어 적은 부호화기로 요구하는 상황에 적합한 여러 조합의 영상을 손쉽게 추출하여 생성해낼 수 있다. 또한 중심 영역에 가중치를 주는 방식을 통해 스트리밍 시나리오에서 비트레이트 절약을 개선시켰으며, 사용자 시점을 토대로 품질 비교를 진행하여 실제 사용

자의 경험을 평가하고자 시도하였다. 향후 스트리밍 환경에서 중요도 예측 맵 뿐만이 아닌 미리 예측된 360 영상의 여러 정보들을 토대로, 효율적인 비디오 전송을 위한 연구를 더 진행할 예정이다.

참고 문헌 (References)

- [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] Seehwan Yoo and Eun-Seok Ryu. 2017. Parallel HEVC decoding with asymmetric mobile multicores. Multimedia Tools and Applications 76,

- 16 (2017), 17337 - 17352.
- [3] Hyun-Joon Roh, SungWon Han, and Eun-Seok Ryu. 2017. Prediction complexity based HEVC parallel processing for asymmetric multicores. *Multimedia Tools and Applications* 76, 23 (2017), 25271 - 25284.
- [4] Tuan Thanh Le, Dien Van Nguyen, and Eun-Seok Ryu. 2018. Computing Offloading Over mmWave for Mobile VR: Make 360 Video Streaming Alive. *IEEE Access* (2018).
- [5] Dien Nguyen, Tuan Le, Sangsoo Lee, and Eun-Seok Ryu. 2018. SHVC Tile-Based 360-Degree Video Streaming for Mobile VR: PC Offloading Over mmWave. *Sensors* 18, 11 (2018), 3728.
- [6] Stefano Petrangeli, Viswanathan Swaminathan, Mohammad Hosseini, and Filip Turck. (2017). An HTTP/2-Based Adaptive Streaming Framework for 360° Virtual Reality Videos. In *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*. ACM, New York, NY, USA, 306 - 314.
- [7] Cagri Ozcinar, Ana De Abreu, and Aljoscha Smolic. "Viewport-aware adaptive 360° video streaming using tiles for virtual reality." 2017 IEEE International Conference on Image Processing (ICIP) (2017): 2174-2178.
- [8] Jang-Woo Son, Dongmin Jang, Eun-Seok Ryu. 2018. Implementing Motion-Constrained Tile and Viewport Extraction for VR Streaming. *ACM Network and Operating System Support for Digital Audio and Video 2018 (NOSSDAV2018)*.
- [9] Robert Skupin, Yago Sanchez, Karsten SÄijhring, Thomas Schierl, Eun-Seok Ryu, and Jangwoo Son. 2018. Temporal MCTS Coding Constraints Implementation. 122th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 122/m42423.
- [10] Dongmin Jang, Jang-Woo Son, JongBeom Jeong, Eun-Seok Ryu, "Implementing Renderer for Viewport Dependent 360 Video", *Journal of Broadcast Engineering (JBE)* Vol. 23, No. 6, Nov. 2018.
- [11] Jang-Woo Son, Dongmin Jang, JongBeom Jeong, Eun-Seok Ryu, "Viewport-Based 360 Degree Video Streaming using Motion-Constrained Tile Set", *The Korean Institute of Broadcast and Media Engineers (KIBME) Summer Conference*, pp.092-095, Jun. 21-23, 2018.
- [12] Robert Skupin, Yago Sanchez, Dimitri Podborski, Cornelius Hellge, and Thomas Schierl, "Viewport-dependent 360 degree video streaming based on the emerging Omnidirectional Media Format (OMAF) standard," *Image Processing (ICIP)*, 2017 IEEE International Conference on. IEEE, Beijing, China, pp. 4592-4592, 2017.
- [13] Matthias Kümmerer, Lucas Thei and Matthias Bethge, (2014). "Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet."
- [14] Laurent itti, Christof Koch and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [15] Borji, Ali. (2018). "Saliency Prediction in the Deep Learning Era: An Empirical Investigation."
- [16] Junting Pan, Cristian Canton-Ferrer, Kevin McGuinness, Noel E. O'Connor, Jordi Torres, Elisa Sayrol and Xavier Giró. "SalGAN: Visual Saliency Prediction with Generative Adversarial Networks." *ArXiv abs/1701.01081* (2017).
- [17] Jesús Gutiérrez, Erwan 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", *International Conference on Quality of Multimedia Experience (QoMEX)*, Sardinia, Italy, May. 2018.
- [18] Mikhail Startsev and Michael Dorr, "360-aware saliency estimation with conventional image saliency predictors.", *Signal Processing: Image Communication* 69 (2018), 43 - 52, 2018.
- [19] Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara, "Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model", *IEEE Transactions on Image Processing* 27, 10 (Oct 2018), 5142 - 5154.
- [20] Jonathan Harel, Christof Koch and Pietro Perona, "Graph-Based Visual Saliency." *Adv. Neural Inform. Process. Syst.* 19. 545-552, 2006.
- [21] Eleonora Vig, Michael Dorr and David Cox, "Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2798-2805. 10.1109/CVPR.2014.358, 2014.
- [22] Mohammad Hosseini, Viswanathan Swaminathan. "Adaptive 360 VR Video Streaming: Divide and Conquer." In *2016 IEEE International Symposium on Multimedia (ISM)*. 107 - 110, 2016.
- [23] Ramin Ghaznavi-Youvalari, Alireza Zare, Alireza Aminlou, Miska M. Hannuksela and Moncef Gabbouj, "Shared Coded Picture Technique for Tile-Based Viewport-Adaptive Streaming of Omnidirectional Video," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3106-3120, Oct. 2019.
- [24] Yashas Rai, Patrick Le Callet, and Philippe Guillotel. 2017. "Which saliency weighting for omnidirectional image quality assessment?," In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. 1 - 6, 2017.
- [25] Heinrich Hertz Institute Fraunhofer Institute for Telecommunications. 2018. High Efficiency Video Coding (HEVC) reference software HM. <https://hevc.hhi.fraunhofer.de/>.

저 자 소 개



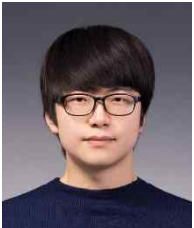
이 순 빈

- 2014년 3월 ~ 현재 : 가천대학교 컴퓨터공학과 학사과정
- ORCID : <https://orcid.org/0000-0002-8951-0335>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



장 동 민

- 2019년 2월 : 가천대학교 컴퓨터공학과 학사
- 2019년 3월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 석사과정
- ORCID : <https://orcid.org/0000-0001-7920-8637>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



정 종 범

- 2018년 8월 : 가천대학교 컴퓨터공학과 학사
- 2018년 9월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 석사과정
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



이 상 순

- 1982년 2월 : 인하대학교 전자공학과 학사
- 1986년 2월 : 인하대학교 전자계산학과 석사
- 2005년 2월 : 인천대학교 컴퓨터공학과 박사
- 1994년 2월 ~ 현재 : 가천대학교 컴퓨터공학과 부교수
- ORCID : <https://orcid.org/0000-0001-6680-2637>
- 주관심분야 : 컴퓨터네트워크, 시스템소프트웨어, IoT



류 은 석

- 1999년 8월 : 고려대학교 컴퓨터학과 학사
- 2001년 8월 : 고려대학교 컴퓨터학과 석사
- 2008년 2월 : 고려대학교 컴퓨터학과 박사
- 2008년 3월 ~ 2008년 8월 : 고려대학교 연구교수
- 2008년 9월 ~ 2010년 12월 : 조지아공대 박사후과정
- 2011년 1월 ~ 2014년 2월 : InterDigital Labs Staff Engineer
- 2014년 3월 ~ 2015년 2월 : 삼성전자 수석연구원/파트장
- 2015년 3월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 조교수
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 조교수
- ORCID : <https://orcid.org/0000-0003-4894-6105>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 코딩 및 국제 표준, HMD/VR 응용분야