

일반논문 (Regular Paper)

방송공학회논문지 제27권 제2호, 2022년 3월 (JBE Vol.27, No.2, March 2022)

<https://doi.org/10.5909/JBE.2022.27.2.216>

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

6DoF 몰입형 비디오 스트리밍을 위한 그룹 분할 기반 적응적 렌더링 기법

이 순 빈^{a)}, 정 중 범^{a)}, 류 은 석^{a)†}

Group-based Adaptive Rendering for 6DoF Immersive Video Streaming

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

요 약

MPEG-I (Immersive) 그룹에서는 6자유도(DoF: degrees of freedom)를 제공하는 몰입형 비디오의 표준화 프로젝트를 진행 중에 있다. MPEG Immersive Video (MIV) 표준화 기술에서는 사용자에게 움직임 시차(parallax)를 제공하기 위해 취득한 다수의 영상을 깊이 맵 기반 이미지 렌더링(depth map-based image rendering, DIBR)을 바탕으로 임의의 사용자 시점의 뷰를 렌더링하게 된다. 현재 MIV에서는 효율적인 부호화를 위한 기술들이 많이 논의된 바 있지만, 전송 측면에 대해서는 여전히 논의가 필요하다. 본 논문은 사용자 시점에 적응적인 몰입형 비디오 스트리밍을 위한 품질 할당 기법을 제안한다. 현재 MIV에서 지원하고 있는 그룹 분할 기법을 통하여 독립적으로 전송, 복원이 가능한 시점 그룹 단위를 생성하여 이를 사용자 시점에 기반한 품질 할당 기법을 통해 효율적인 전송이 가능하도록 한다. 제안하는 적응적 전송 기법은 Test Model for Immersive Video (TMIV) 시험모델을 통해 구현되었으며, 주어진 합성 시점 위치에 따라 렌더링 과정에서의 기여도를 그룹별로 계산하고 우선 시점 그룹을 판단하여 고품질로 전송한다. 사용자 시점에 대한 렌더링 비교 결과를 통해 제안하는 기법이 기존 TMIV 대비 PSNR 지표에서 평균 17.0%, IV-PSNR 지표에서 14.6%의 BD-rate 감소율을 보여 효율적인 전송이 가능함을 보였다.

Abstract

The MPEG-I (Immersive) group is working on a standardization project for immersive video that provides 6 degrees of freedom (6DoF). The MPEG Immersion Video (MIV) standard technology is intended to provide limited 6DoF based on depth map-based image rendering (DIBR) technique. Many efficient coding methods have been suggested for MIV, but efficient transmission strategies have received little attention in MPEG-I. This paper proposes group-based adaptive rendering method for immersive video streaming. Each group can be transmitted independently using group-based encoding, enabling adaptive transmission depending on the user's viewport. In the rendering process, the proposed method derives weights of group for view synthesis and allocate high quality bitstream according to a given viewport. The proposed method is implemented through the Test Model for Immersive Video (TMIV) test model. The proposed method demonstrates 17.0% Bjontegaard-delta rate (BD-rate) savings on the peak signal-to-noise ratio (PSNR) and 14.6% on the Immersive Video PSNR(IV-PSNR) in terms of various end-to-end evaluation metrics in the experiment.

Keyword : Immersive Video, MPEG-I, Adaptive Streaming, TMIV, 6DoF

Copyright © 2022 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

사용자에게 보다 높은 시점 자유도(Degrees of Freedom, DoF)를 제공하려는 차세대 미디어 기술에 대한 요구사항은 계속해서 증대되고 있다. 2018년, MPEG-I 프로젝트(ISO/IEC 23090)에서는 전방위 미디어 포맷(Part 2 Omnidirectional Media Format, OMAF)을 통해 3차원 공간에서 감상자의 머리 방향에 따른 세 가지 자유도를 제공하는 360 VR 영상을 3DoF로 정의하여 표준화를 진행하였다^[1]. 이러한 3DoF 기술의 발전과 더불어 보다 몰입감 있는 시청이 가능하도록 3차원 공간에서의 움직임에 대한 자유로운 시점 변환을 제공하는 6DoF 비디오 기술에 대한 관심 또한 증가하였으며, MPEG 125차 회의에서는 점진적으로 최대 6DoF를 지원하는 미디어 시스템 기술에 대한 몰입형 비디오(Part 12: MPEG Immersive video, MIV)를 정의하고 표준화를 시작하였다^[2-3].

MIV 표준화 기술은 다양한 위치에서 획득한 다시점 영상을 기반으로, 시점에 따른 영상 합성을 통해 사용자에게 보다 높은 자유도와 몰입감 있는 경험을 제공하고자 한다. 당초 MIV는 3DoF+, 제한적인 범위 내에서 화면 전환 이동에 따른 운동 시차(motion parallax)를 제공하는 것을 목적으로 하여, 2019년부터 3DoF+ 비디오의 표준화를 본격적으로 시작하고 더 나아가 6DoF를 제공하기 위해 6DoF의 EE (Exploration Experiment)를 진행하고 있다^[4-6]. 2019년 3월 MPEG 126차 회의에서는 MPEG-I (Immersive)에서는 몰입형 비디오 표준화 기술을 지원하는 테스트 모델(Test Model for Immersive Video, TMIV)를 정의하였으며, 이후 MIV 기술은 129차 회의에서 비디오 기반 포인트 클라우드 압축(Part 5: Video-based Point Cloud Compression,

V-PCC)과의 유사성을 고려하여 용어, 신택스 구문, 복호화 과정 등 통합을 논의하였고 그 결과 130차 회의에서 공통 명세인 Visual Volumetric Video-based Coding (V3C)을 정의하였다. V3C는 V-PCC와 MIV의 확장 메커니즘을 제공하며, MIV 역시 V-PCC 표준 기술과 함께 V3C 비트스트림 구조를 갖추게 되었다^[7].

MIV의 기술적 배경은 복수 개의 다시점 비디오를 전송하고 이를 감상하고자 할 때, 압축 및 복원 과정에서 비디오 코덱 요구사항을 최소화하도록 데에 있다. 다시점 비디오 압축을 진행하기 전, TMIV를 통하여 전처리 과정으로 각 시점 간의 중복성을 제거하여 차분 패치만을 모아둔 아틀라스(Atlas) 영상으로 재구성한다. 이렇게 재구성된 아틀라스 영상은 비디오 코덱 요구사항에서 픽셀 레이트(pixel rate)를 크게 절감하는 것으로 필요한 디코더 인스턴스를 최소화하여 비디오 코덱에 대한 하드웨어 제약을 크게 완화할 수 있다^[22,25]. 공통 실험 조건에 명시된 HEVC Main 10의 5.2 Profile 레벨을 기준으로, 하나의 디코더 인스턴스는 최대 4096x2048까지의 해상도 영상을 지원하며 그 이상의 해상도일 경우에는 추가적인 디코더 인스턴스가 필요하다. 특히 최적화된 병렬처리 성능을 내기 어려운 모바일 등의 단말에서는 디코더 인스턴스에 개수에 따른 성능 영향이 크게 나타나며, 따라서 MIV 표준화 기술은 비디오 코덱에서 처리해야 하는 영상 해상도를 시점 간 중복성 제거를 통해 절감하고자 한다^[18].

그러나 영상 해상도를 절감하기 위해 시점 간의 중복성 제거 과정을 거칠 경우, 복원 과정에서 서로의 시점들이 의존성을 가지게 되어 독립적인 단위의 복원과 전송이 어려워지는 문제로 기존의 적응적 스트리밍 기법이 적용되기 어려운 문제를 갖고 있다. 본 논문에서는 TMIV의 다시점 처리 과정에서 독립적인 전송과 복원을 보장하는 그룹 기반 TMIV를 기반으로, 사용자 시점에 따라 각 그룹의 렌더링 기여도를 계산하여 시점 그룹의 품질을 적응적으로 전송하는 방법을 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안하는 방법의 기반이 되는 그룹 분할 기반 인코딩에 대한 전반적인 내용을 설명하고, 3장에서는 제안하는 기법에 대해 서술한다. 4장에서는 실험 환경 서술과 성능 평가 결과를 분석하며, 마지막으로 5장에서는 결론을 맺는다.

a) 성균관대학교 컴퓨터교육학과(Department of Computer Science Education, Sungkyunkwan University)

‡ Corresponding Author : 류은석(Eun-Seok Ryu)
E-mail: esryu@skku.edu
Tel: +82-2-760-0677
ORCID: <http://orcid.org/0000-0003-4894-6105>

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-00920, Development of ultra high resolution unstructured plenoptic video storage/compression/streaming technology for medium to large space).

· Manuscript received January 27, 2022; Revised March 18, 2022; Accepted March 18, 2022.

II. 관련 연구

TMIV에서 모든 시점을 한번에 고려하여 중복성 제거 과정을 시행하게 될 경우, 깊이맵에서의 전경과 배경 구분이 어려운 문제로 최종 렌더링 과정에서 배경에 해당하는 패치가 전경 물체에 합성되는 등의 합성 오류를 일으킬 수 있다는 문제가 지적되었다. 따라서 이러한 사항을 개선하기 위해 시점을 단위 별로 나누어 서로 다른 그룹으로 처리하는 그룹 기반 기법은, 렌더링된 합성 영상 화질 평가에서 개선된 결과를 보고하여 MPEG 128차 회의에서 채택된 기술이다⁸⁾. 일반적으로 그룹 기반 기법의 압축 효율은 깊이맵이 정확하지 못한 자연 영상일수록, 하나의 시점이 넓은 시청공간을 커버하지 못하는 perspective 영상일수록 이득이 더 뛰어나며(최대 43%의 BD-rate 이득), 반대로 equi-rectangular projection (ERP) 영상 콘텐츠에 대해서는 시점의 분할이 중복성을 높이기 때문에 오히려 큰 손실을 나타내는 경향을 보인다⁹⁾. 예를 들어, ERP 영상인 Museum 시퀀스에서는 하나의 시점만으로 대부분의 시청 영역을 표현할 수 있어 기본적으로 바탕 시점이 1개로 선택되지만, 그룹 분할 기법에서는 그룹마다 각각의 바탕 시점을 가지게 되어 불필요한 시점의 중복성이 커지게 된다²³⁾.

그룹 기반 기법의 기술적 배경은 렌더링 화질 개선에 있지만, 해당 기법을 통해 나뉘어진 그룹들은 서로 독립적으로 TMIV의 처리 과정을 거쳐 그룹 간의 의존성이 존재하지 않는 특징을 가지고 있다. 본 논문에서는 이러한 그룹 기반 기법이 별개의 영상 단위로 처리되어 독립적으로 복

원, 전송이 가능하다는 측면에서 적응적 스트리밍에 활용될 수 있는 새로운 기법을 제안한다. 대표적인 적응적 스트리밍 기법의 예시로 360 비디오에서의 타일(Tile) 기반 스트리밍을 기반으로 한 Region of Interest (RoI) 기반 전송 기법 등을 들 수 있으며, V-PCC에서도 타일 기반 방법에 대한 논의가 진행되고 있다¹⁰⁻¹³⁾. 이와 같이 차세대 미디어의 주요 기능 중 하나인 공간적 임의 접근(spatial random access)을 고려한 논의가 활발하게 진행되고 있지만 몰입형 비디오 기술에서는 타일 기반 접근 방법 이외에 아직 논의가 활발하게 이루어지지 않고 있다¹⁴⁻¹⁵⁾.

그림 1은 그룹 분할 방법을 통한 그룹 기반 TMIV 인코딩의 인코더 블록 다이어그램을 나타낸다. 그룹 기반 인코딩의 구체적인 동작은 다음과 같이 이루어진다. 1) 생성될 아틀라스의 최대 크기와 수를 파라미터로 입력받아 추후 각 그룹의 파라미터 선택에 포함한다. 2) 입력받은 그룹 개수 파라미터를 따라 각 그룹을 분할한다. 이 때, 입력받은 카메라 파라미터를 이용하여 3차원 축 중에서 가장 넓은 범위를 가지는 축을 중심 축으로 설정하고, 중심 축에 해당하는 영상들을 파라미터 제한을 넘지 않는 선에서 우선적으로 그룹으로 포함한다. 3) 분할된 그룹의 영상들은 각각 독립적으로 TMIV 처리 과정인 프루닝, 패킹, 클러스터링 등의 과정을 거쳐 아틀라스로 변환된다. 이렇게 생성된 각 그룹의 아틀라스들은 그룹 간 의존성이 없으므로 독립적으로 복원, 전송될 수 있다. 또한 카메라 축을 기준으로 서로 다른 그룹으로 분할되기 때문에, 시점 렌더링 측면에서도 각 그룹은 서로 다른 방향의 시점을 포함하고 있는 경향을 나타낸다.

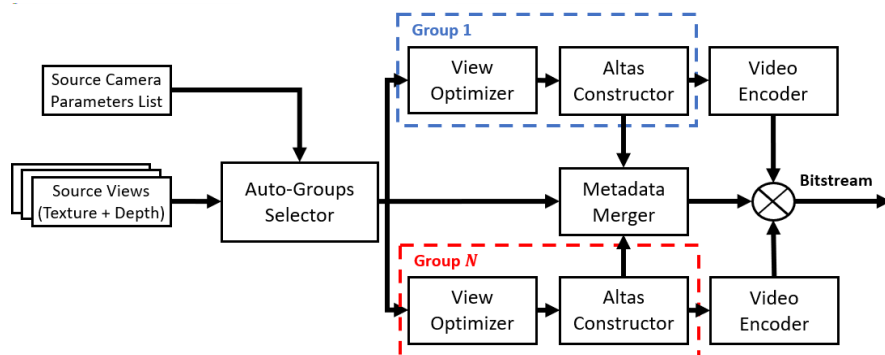


그림 1. 그룹 기반 TMIV 인코더의 블록 다이어그램
 Fig. 1. Block diagram of Group-based TMIV Encoder

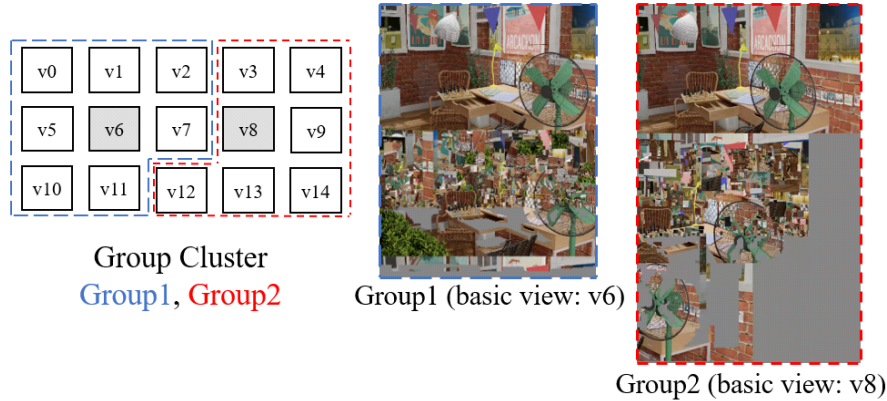


그림 2. Fan 시퀀스의 그룹 분할 기법 적용 결과 예시
 Fig. 2. Example of group-based encoding atlases (Fan Sequence)

그림 2는 Fan 영상에서 그룹 분할 기법을 적용한 결과를 나타낸다. 각 그룹의 아틀라스는 기존 TMIV에서 처리되던 것과 같이, 바탕 시점과 추가 시점의 패치들을 독립적으로 생성하게 된다. 그림 2에서는 평면 배열로 15개의 시점이 존재하는 다시점 영상에 대해 2개의 그룹으로 분할된 결과이다. 이렇게 생성된 아틀라스를 이용해 사용자 시점을 렌더링하고자 할 때, 정해진 대역폭에서 각 그룹에 품질을 비례적으로 할당하는 경우를 생각해볼 수 있다. 이 경우, 사용자 가상 시점 렌더링에 높은 기여도를 가지는 그룹에 대하여 보다 높은 품질로 전송하는 것이 효율적이라고 판단할 수 있다. 본 논문에서는 가상 시점에 대한 그룹 별 기여도를 TMIV의 렌더러에서 먼저 계산하고, 이 정보를 전송시에 활용하는 방법을 제안한다. 또한 그룹 분할 기법을 기반으로 사용자 시점에 대한 렌더링 실험과 평가를 통해 제안하는 적응적 렌더링 기법이 효율적인 전송이 가능함을 보인다.

III. 제안 기법

현재 TMIV에서 가상 시점을 합성하여 렌더링하기 위한 렌더러 소프트웨어로 VWS(View Weighting Synthesizer)를 채택하고 있다¹⁶⁾. VWS의 주요 특징은 주변 영상을 이용해 가상 시점 영상을 합성하고자 할 때, 가시도(Visibility)를 계산하여 이를 렌더링 과정에서 반영한다는 점에 있

다. VWS에서 합성하고자 하는 사용자 시점의 위치를 입력받게 되면 하나의 깊이맵을 생성하고 이를 입력 영상들에 와핑(warping)을 수행하여 가시도 맵(Visibility map)을 생성한다. 이렇게 생성한 가시도 맵을 이용해 픽셀 단위로 주변 영상의 기여도를 계산하고, 렌더링 과정에서 이 정보를 이용하여 보다 정확하고 높은 품질의 가상 시점 합성이 가능하다²⁴⁾.

본 논문에서는 VWS의 렌더링 방법에 착안하여 입력받은 시점에 따라 각 그룹 영상들의 기여도를 사전에 계산하고, 이를 우선 시점 그룹 전송 시에 반영할 수 있도록 하는 방법을 제안한다. 기여도가 높은 영상이 많이 포함되어 있는 그룹을 우선 시점 그룹으로 정의하여, 적응적 렌더링 프레임워크에서 우선 시점 그룹에 대해 높은 품질을 적용한 비트스트림을 선택하여 가상 시점을 렌더링하는 방법을 거치게 된다. 즉, 영상 전송 과정에서 입력받은 사용자 시점을 렌더링하고자 하였을 때의 주변 영상에서의 기여도를 비트스트림 품질 선택 시에 활용한다.

그림 3은 Group 시퀀스에서의 그룹 분할에 따른 카메라 좌표와 사용자 시점 기록 데이터셋을 활용한 시점(Pose trace)의 첫 프레임의 가상 시점 좌표를 나타낸다. 앞서 기술한 그룹 분할 알고리즘에 따라 각 카메라 시점이 그룹으로 분산되어 분할된 것을 확인할 수 있으며, 가상 시점은 가까운 그룹에 있는 시점들을 주로 활용하여 시점을 합성하게 될 것을 관찰할 수 있다. 예를 들어, p01과 p02는 주로 그룹 1의 시점을 활용하여 시점 합성을, p03의 경우 복합적

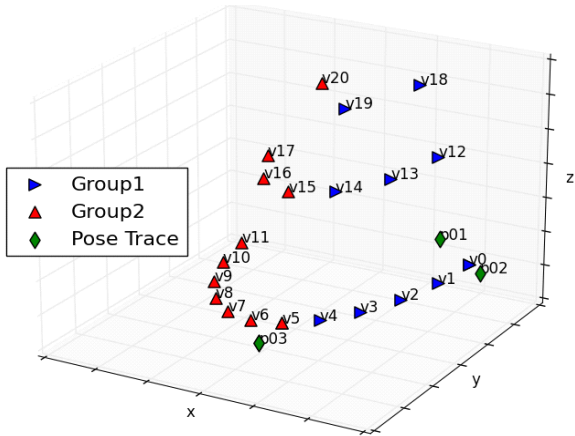


그림 3. Group 시퀀스의 각 시점 카메라 좌표 정보
 Fig. 3. Camera coordinate information for each viewpoint (Group sequence.)

으로 시점을 활용하게 된다. 따라서 동일한 대역폭이라 하더라도 사용자 시점에 기여도가 높은 시점 그룹은 우선적으로 높은 품질의 비트스트림을 전송하는 것으로 몰입형 비디오에서의 효율적인 적응적 스트리밍이 가능할 것이다. 본 논문에서 제안하는 방법을 검증하기 위해, 스트리밍

시나리오를 가정하여 타겟 비트레이트를 설정하고 앞서 기술한 방법을 적응적 렌더링 프레임워크에 활용한다. 특정 타겟 비트레이트와 사용자 시점에 대한 타겟 좌표가 주어졌을 때, 각 영상의 시점 기여도 수치를 계산하여 우선 시점 그룹을 선택한다. 이후, 선택된 우선 시점 그룹에 대하여 먼저 우선적으로 높은 품질의 비트스트림을 선택하여 비트레이트를 할당한다. 이후 우선 시점 그룹이 아닌 그룹은 나머지 비트레이트에 대응되는 품질 비트스트림을 선택하여 이를 전송하는 과정을 거치게 된다. 우선 시점 그룹에 대한 비트레이트 할당 비율은 시점 기여도 수치에 따라 변화될 수 있으며, 4장에서 실험 결과를 보인다. 제안하는 적응적 렌더링 프레임워크를 통해 가상 시점을 렌더링하는 데에 기여도가 큰 우선 시점 그룹이 높은 품질로 전송되므로, 각 그룹의 품질을 균일하게 전송할 때보다 효율적인 적응적 전송이 가능하다.

그림 4는 제안하는 적응적 전송 프레임워크를 나타낸다. 앞서 서술한 그룹 분할 기법을 통해 분할된 그룹 아틀라스들을 다수의 양자화 파라미터(Quantization Parameter, QP) 값으로 인코딩하여 비트스트림 형태로 저장한다. 다음으로

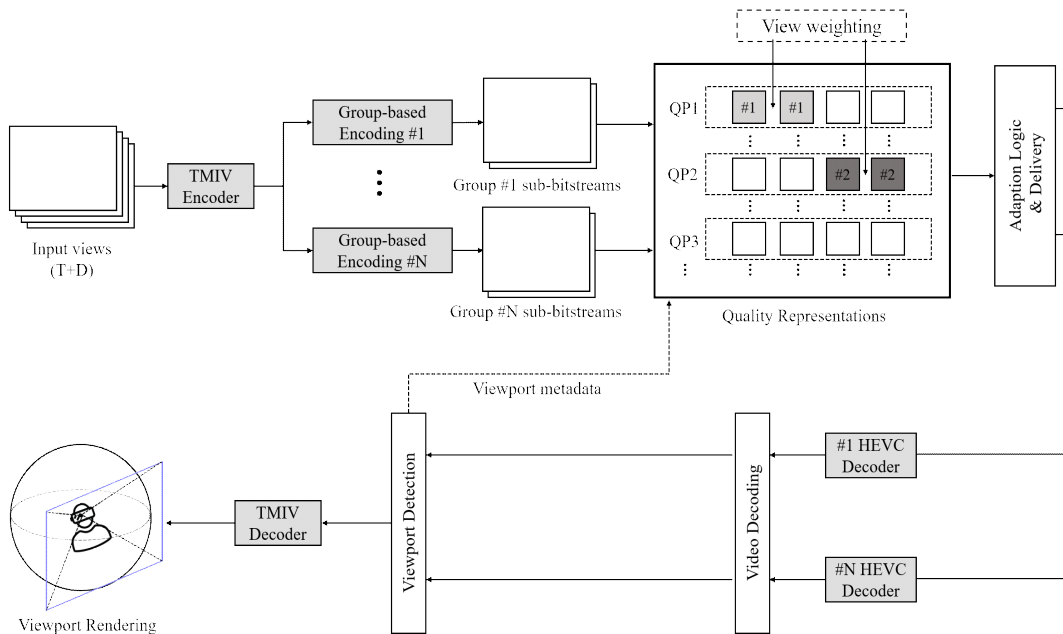


그림 4. 본 논문에서 제안하는 적응적 몰입형 비디오 전송 프레임워크
 Fig. 4. Proposed adaptive immersive video rendering framework

정해진 대역폭과 렌더링할 사용자 시점의 정보를 입력받게 되면 렌더러에서 시점 별 기여도를 사전에 계산하고, 이 ‘view weighting’ 정보를 통해 대역폭에 따라 각 그룹의 비트스트림 품질을 선택한다. 본 논문에서의 비트스트림 품질 선택 방법의 경우, 전송 프레임워크의 기본적인 검증을 위해 비우선 시점 그룹에 대해 $\Delta QP+5$ 를 적용하는 방법을 선택하였다. 즉, 우선 시점 그룹이 아닌 그룹의 경우 우선 시점 그룹보다 양자화 파라미터를 5만큼 높은 비트스트림을 전송받아 렌더링하며, 추후 기여도 정보를 이용한 최적의 여러 적응적 알고리즘이 포함될 수 있다.

IV. 실험 결과 및 분석

제안하는 실험 방법은 MPEG-I의 테스트 모델인 TMIV 6.0과 함께, 공통 실험 조건(Common Test Conditions, CTC)을 준수하여 진행하였다^[17-18]. 테스트 콘텐츠로는 Museum (SB), Painter (SD), Frog (SE), Fan (SO), Group (SR), Carpark (SP), Street (SU), Hall (ST) 8가지 시퀀스를 선정하였다. 타겟 비트레이트는 영상 콘텐츠에 따라 다른 비트레이트를 가지며, 깊이 맵의 비트레이트는 텍스처 양자화 파라미터의 선형 변환 식이 제시된 공통 실험 조건을 준수하였다.

현재 MPEG-I의 공통 실험 조건에서는 가상 시점에 대해 합성된 영상의 품질을 관찰하기 위하여, 각 테스트 시퀀스마다 실제 감상자의 시선 좌표를 기록한 pose trace 데이터

셋 또한 함께 포함하고 있다. 본 논문에서도 해당 pose trace 데이터 셋을 통해 렌더링된 가상 시점에 대한 화질 측정을 통해 제안하는 방법의 효율성을 검증한다. Pose trace를 통하여 렌더링된 가상 시점에 대한 화질 측정은 원본 영상을 이용해 렌더링된 시점이라 하더라도, 합성 오류가 여전히 발생할 수 있기 때문에 현재까지도 객관적 평가에 있어서 어려움이 있다. MPEG 회의에서도 주관적 화질 평가 등 다양한 탐색 실험과 논의를 진행 중에 있으며, 본 논문에서는 제공하는 3가지 pose trace 데이터 모두의 경우에 대해 다양하게 객관적 및 주관적 화질 평가 결과를 제시함으로써 제안하는 기법의 우수성을 보인다^[19-20].

표 1은 TMIV에서 그룹 분할 기반 시 선택된 시점과 카메라 배열 특성을 나타낸다. 그룹 분할 기법은 시점 그룹 간 서로의 의존성을 제거하는 만큼, 기존 TMIV 대비 영상 간의 중복성이 늘어날 수 있는 단점을 가지고 있으며 그에 따라 콘텐츠 별로 압축 효율이 다양하게 나타난다. 앞서 관련 연구 파트에서 서술했듯이 일반적으로 하나의 시점이 큰 시야각을 가지는 ERP 영상 콘텐츠의 경우, 하나의 시점으로도 많은 가상 시점을 표현할 수 있기 때문에 그룹 분할 기반 시 압축 효율에 있어서 손실을 보인다. 반대로 각 시점이 커버하는 영역이 작으나 전체 가상 시점이 큰 범위를 표현해야 하는 Perspective 콘텐츠의 경우 그룹 분할 기반 방법이 압축 효율에 이득을 보이는 경향을 나타낸다.

공통 실험 조건 영상에서 ERP 콘텐츠의 대표적인 예시로 Museum (SB) 시퀀스를 들 수 있으며, Perspective 콘텐츠

표 1. 테스트 시퀀스의 특성
 Table 1. Properties of test video sequences

Sequence name	Class	No. of source view	Resolution	Category	View FoV
Museum	B	24	2048x2048	Computer-generated	180° × 180° ERP
Painter	D	16 (4x4)	2048x1088	Natural contents	50° × 37°
Frog	E	13 (13x1)	1920x1080	Natural contents	63.65° × 38.47°
Group	R	21	1920x1080	Computer-generated	75° × 48°
Fan	O	15 (5x3)	2048x2048	Computer-generated	50° × 50°
Carpark	P	9 (9x1)	1920x1088	Natural contents	63° × 48°
Street	U	9 (9x1)	1920x1088	Natural contents	63° × 48°
Hall	T	9 (9x1)	1920x1088	Natural contents	63° × 48°

표 2. 영상 별 그룹 분할 결과 및 렌더링 시점에 대한 그룹 기여도 결과 (평균화된 결과)
 Table 2. Results of group-based encoding and view weights (normalized)

Class	Views per group	p01 view weights	p02 view weights	p03 view weights
B	v1, v4, v2, v8, v18, v7, v17, v9, v6, v22, v11, v21 / v5, v15, v13, v16, v10, v12, v0, v3, v20, v19, v14, v23	G1(51.84%) G2(49.26%)	G1(44.80%) G2(55.20%)	G1(52.14%) G2(47.86%)
D	v0, v4, v1, v5, v2, v8, v6, v9 / v10, v3, v12, v13, v7, 11, v14, v15	G1(64.95%) G2(35.05%)	G1(81.00%) G2(19.00%)	G1(76.19%) G2(23.81%)
E	v0, v1, v2, v3, v4, v5 / v6, v7, v8, v9, v10, v11, v12	G1(98.07%) G2(1.93%)	G1(10.64%) G2(89.36%)	G1(7.18%) G2(92.82%)
R	v0, v1, v12, v2, v13, v3, v18, v14, v4, v19 / v5, v15, v6, v20, v16, v7, v17, v8, v9, v10, v11	G1(96.28%) G2(3.72%)	G1(96.17%) G2(3.83%)	G1(10.51%) G2(89.49%)
O	v14, v9, v13, v8, v4, v12, v3 / v7, v2, v11, v6, v1, v10, v5, v0	G1(80.99%) G2(19.01%)	G1(36.14%) G2(63.86%)	G1(55.88%) G2(44.12%)
P	v8, v7, v6, v5 / v4, v3, v2, v1, v0	G1(86.77%) G2(13.23%)	G1(20.99%) G2(79.01%)	G1(90.44%) G2(9.56%)
U	v8, v7, v6, v5 / v4, v3, v2, v1, v0	G1(92.55%) G2(19.01%)	G1(93.53%) G2(19.01%)	G1(95.22%) G2(19.01%)
T	v8, v7, v6, v5 / v4, v3, v2, v1, v0	G1(84.32%) G2(15.68%)	G1(90.97%) G2(9.03%)	G1(10.99%) G2(89.01%)

츠에서는 Frog (SE) 시퀀스를 예로 들 수 있다. 컴퓨터 그래픽을 통해 생성된 영상이 아닌 실제 취득한 자연 영상은 장비 상의 한계로 깊이 맵에서 보다 부정확함을 나타내며, 따라서 그룹 분할 시에 시점 분리에 따른 합성 오류 개선 또한 효과가 크게 나타난다. 제안하는 적응적 전송 알고리즘은 이러한 콘텐츠의 특성에 상관없이 성능 개선을 나타냄을 4장에서 보인다.

표 2는 공통 실험 조건 상의 Pose Trace별 첫 프레임에서 우선 시점 그룹의 기여도와 채택 여부를 나타낸다. 각 사용자 시점 p01, p02, p03 시점을 렌더링 시의 기여도를 모든 영상에 대하여 계산한 후, 이를 그룹 별로 합산하여 나타내었다. 그룹 분할 개수는 모든 콘텐츠에 대하여 N=2로 실험하였으며, 합산 기여도가 높은 그룹을 우선 시점 그룹으로 정의하고 3장에서 기술한 적응적 프레임워크를 적용하여 실험 결과를 나타낸다. 타겟 비트레이트는 각 콘텐츠의 비트레이트 매칭 문제로 인해 콘텐츠마다 별도로 임의로 구간을 제한하였다.

그림 5는 p02 시점에서 렌더링된 그룹 분할 기반 방법과 제안하는 알고리즘과의 BD-rate 결과를 그래프로 나타낸

다. 제안하는 방법은 그룹 기반 방법과 객관적 화질 수치를 비슷하게 유지하면서도 전송에 필요한 비트레이트가 크게 줄어든 경향을 확인할 수 있다. 상대적으로 시점 합성에 기여도가 낮은 그룹에 대한 품질이 감소하였으므로, 비트레이트를 크게 절약하면서도 합성 시점에 대한 품질이 크게 떨어지지 않는 것을 관찰할 수 있다.

그림 6은 ‘p03’ 시점에서 제안하는 알고리즘의 주관적 화질 비교 결과를 나타낸다. 각각 Frog (SE) 시퀀스에서 15000kbps를 타겟 비트레이트로, Painter (SD) 시퀀스에서는 3000kbps를 타겟 비트레이트로 렌더링하였을 때의 비교 결과 이미지를 나타낸다. 그룹 분할 기반 결과는 기존 표준 문서에서 보고된 것과 같이 렌더링 시에 시점 합성 오류를 개선하는 모습을 보인다⁹⁾. 표시된 부분은 제안하는 적응적 품질 할당 기법을 적용한 결과에서 화질 개선이 식별되는 부분을 나타낸다. 제안하는 기법은 합성 우선 시점 그룹에 대해 높은 품질을 할당함으로써 동일 비트레이트 내에서 주로 텍스처 부분에 대한 패치가 고품질로 대체되어 선명하게 나타나며, 객관적 화질 비교 또한 향상된 것을 관찰할 수 있다.

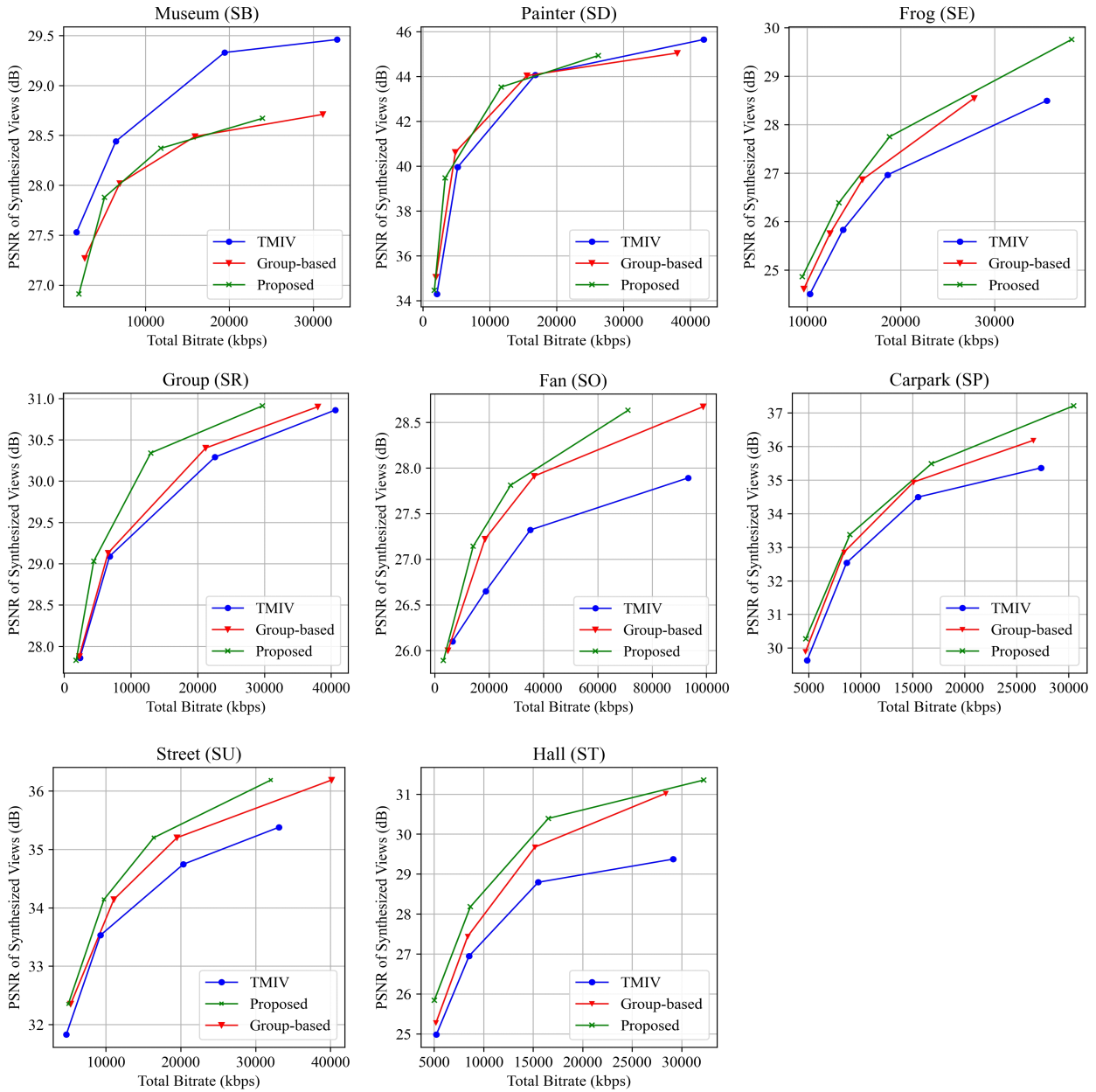


그림 5. 'p02' 시점에 대한 객관적 렌더링 화질 비교 그래프 (TMIV, 그룹 분할 기반, 제안 방법)
 Fig. 5. Objective performance comparison with synthesized 'p02' BD-rate (TMIV, Group-based, Proposed)

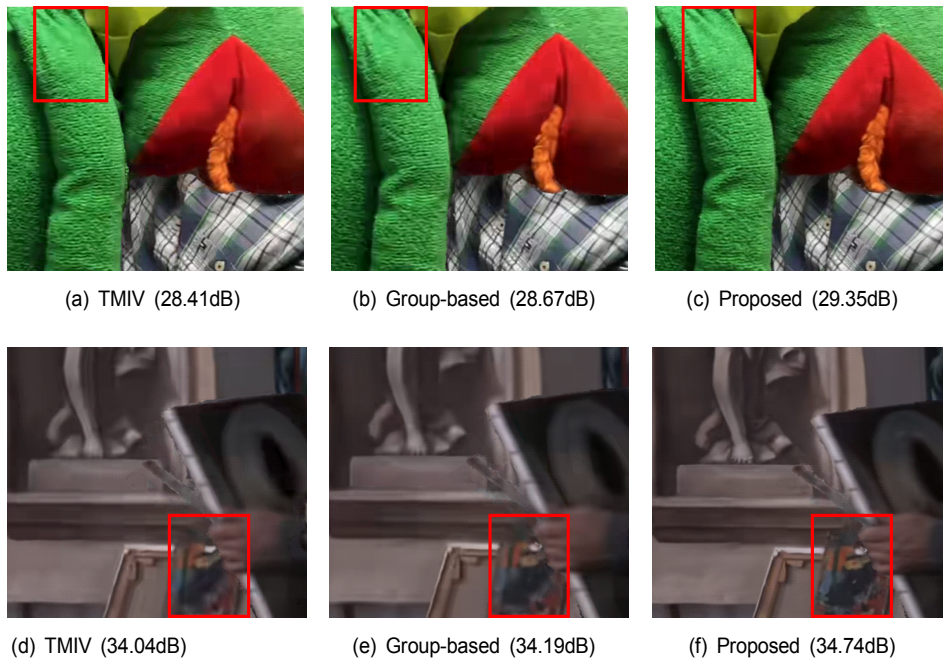


그림 6. 'p03' 시점에 대한 주관적 렌더링 화질 비교 결과 (TMIV, 그룹 분할, 제안 방법)

Fig. 6. Subjective performance comparison with 'p03' views (TMIV, Group-based, Proposed) (a), (b), (c) : Frog (SE), target bitrate 15000kbps, (d), (e), (f) : Painter (SD), target bitrate 3000kbps

표 3. 각 렌더링 시점에 대한 객관적 화질 비교 결과 (TMIV 대비 Group-based, 제안 방법, PSNR)

Table 3. Objective performance comparison for pose trace views (TMIV v.s. Group-based, Proposed, PSNR)

BD-rate(%) (v.s.Group-based) (v.s. Proposed)	SB	SD	SE	SR	SO	SP	SU	ST
p01	46.6% 45.3%	-9.7% -13.6%	-18.9% -38.5%	-12.0% -42.9%	-13.6% -31.8%	-11.5% -23.8%	-19.6% -26.7%	-21.4% -30.9%
p02	48.8% 46.5%	-8.8% -12.1%	-22.7% -45.2%	-9.6% -27.3%	-11.5% -23.3%	-10.4% -25.5%	-15.0% -23.3%	-17.7% -26.6%
p03	43.2% 39.7%	-5.4% -9.2%	-21.3% -41.2%	-6.7% -15.2%	-3.5% -8.8%	-13.5% -28.7%	-14.7% -21.5%	-18.2% -25.4%
Avg.	46.2% 43.8%	-7.9% -11.6%	-20.96% -41.6%	-9.4% -28.4%	-9.5% -21.3%	-11.8% -26.0%	-16.4% -23.8%	-19.1% -27.6%
Group v.s. Proposed	-2.6%	-3.7%	-20.64%	-19.0%	-11.8%	-14.2%	-7.4%	-8.5%

표 4. 각 렌더링 시점에 대한 객관적 화질 비교 결과 (TMIV 대비 Group-based, 제안 방법, IV-PSNR)

Table 4. Objective performance comparison for pose trace views (TMIV v.s. Group-based, Proposed, IV-PSNR)

IV BD-rate(%) (v.s.Group-based) (v.s. Proposed)	SB	SD	SE	SR	SO	SP	SU	ST
p01	41.7% 38.8%	-4.5% -8.8%	-14.5% -33.6%	-15.6% -37.8%	-13.5% -30.9%	-8.7% -16.9%	-15.5% -23.2%	-17.8% -26.8%
p02	45.6% 42.2%	-6.7% -10.4%	-18.7% -39.9%	-10.4% -23.6%	-9.8% -22.4%	-8.8% -20.5%	-14.7% -20.6%	-12.5% -23.4%
p03	35.3% 38.6%	2.3% -5.5%	-19.9% -40.2%	-5.2% -14.5%	-2.7% -8.2%	-11.2% -23.5%	-12.8% -19.9%	-11.0% -20.8%
Avg.	40.8% 39.8%	-2.9% -8.2%	-17.7% -37.9%	-10.4% -25.3%	-8.6% -20.5%	-9.5% -20.3%	-14.3% -21.2%	-13.7% -23.6%
Group v.s. Proposed	-1.0%	-5.3%	-20.2%	14.9%	-11.9%	-10.8%	-6.9%	-9.9%

표 3과 표 4는 기존 TMIV 전송 대비 제안하는 방법에 따라 각각 PSNR와 IV-PSNR로 측정된 결과를 나타낸다. 제안하는 방법은 그룹 분할 기법을 기반으로 이루어지므로, 그룹 분할 기법 시의 성능에도 영향을 받기 때문에 두 방법에 대해 모두 비교한 결과를 나타내었다. 표의 첫 줄은 TMIV 대비 그룹 분할 기법에 대한 BD-rate를 나타내며, 두 번째 줄은 TMIV 대비 제안하는 방법에 대한 BD-rate를 나타낸다. 그룹 분할 기법에서 큰 손실을 나타내었던 ERP 영상 콘텐츠는 제안하는 방법에서도 낮은 이득을 나타내었는데, 이는 앞서 서술했듯 하나의 시점마다 시야각이 크기 때문에, 그룹 분할 시에 오히려 중복성이 커지며 이로 인해 표 2에서 나타나듯이 두 그룹의 기여도가 비슷하기 때문에 우선 시점 그룹에 대한 적응적 품질 할당의 의미가 크지 않기 때문이다. ERP 영상 콘텐츠에 대해 그룹 분할 기법이 큰 손실을 나타내므로, 제안하는 방법을 활용하기 위해서는 ERP 영상을 작은 해상도로 분할하는 등 독립적인 시청 영역 단위로 전송할 수 있도록 처리하는 접근이 필요할 것으로 보인다. 제안하는 방법은 1차원 카메라 배열을 가진 자연 취득 영상 Frog (SE) 시퀀스에 대해 가장 큰 이득을 보이며, 넓은 시청 영역과 원형 카메라 배열을 가진 Group (SR) 시퀀스에 대해서도 다른 시퀀스보다 다소 높은 효율을 보인다. 동일한 1차원 카메라 배열이지만 시점 개수와 시청 영역이 Frog (SE) 시퀀스보다 제한된 Carpark (SP), Street (SU), Hall (ST) 시퀀스는 Frog (SE) 시퀀스보다는 낮은 효율을 보인다. 이로 인해 전체 시청 영역이 넓을수록, 그리고 각 시점의 시야각이 작을수록 우선 시점 그룹이 기여도를 지배적으로 차지하여 제안하는 방법이 더 효율적임을 추론할 수 있다.

제안하는 방법은 PSNR 지표에서 기존 TMIV 대비 평균 17.0%, 그룹 분할 기법 대비 평균 10.9%의 BD-rate 감소율을 보였으며, IV-PSNR 지표에서는 기존 TMIV 대비 평균 14.6%, 그룹 분할 기법 대비 평균 10.1%의 BD-rate 감소율을 보여 비슷한 경향을 나타낸다. IV-PSNR은 MPEG-I에서 몰입형 비디오에 대한 품질 평가를 위해 제안되어 주관적 화질 평가에서 더 높은 상관성을 보이는 지표로, 시점 합성 시에 경계 부분에 대한 오류와 전역적인 색상 편차를 고려하는 특성을 가지고 있다^[21]. ERP 영상을 제외한 그룹 분할 시에 이득을 보였던 콘텐츠들은 제안하는 방법에서도 이득

을 보이는 것을 관찰할 수 있으며, 콘텐츠의 카테고리에 대해 큰 차이없이 비슷한 이득 경향을 나타낸다. 따라서 제안하는 방법은 깊이 맵이 비교적 부정확한 자연 취득 영상에서도 효율적인 적응적 전송이 가능함을 확인하였으며, 추후 품질 할당 알고리즘을 최적화한다면 보다 높은 효율의 몰입형 비디오 스트리밍이 가능할 것이다.

V. 결론

본 논문은 몰입형 비디오 전송 과정에서 사용자 시점에 기반한 적응적 스트리밍을 고려한 렌더링 기법을 제안하였다. 기존 360 영상의 적응적 스트리밍 기술은 motion constrained tile set (MCTS) 기법을 토대로 타일(Tile) 단위의 비트스트림 전송과 복원의 독립성을 보장하고 있지만, 몰입형 비디오 기술은 처리 과정에서 다수의 시점 간 의존성이 존재하는 문제로 기존의 적응적 스트리밍 기법을 적용하는 데에 어려움을 가지고 있다. 제안하는 방법은 시점을 그룹 단위로 분할하여 독립성을 보장하는 그룹 분할 기반 방법을 기반으로, 렌더링하고자 하는 사용자 시점에 대해 각 영상의 기여도를 계산한다. 이후, 우선 시점 그룹의 비트스트림의 품질을 고품질로 선택하는 방법을 통해 몰입형 비디오에서의 효율적인 적응적 스트리밍이 가능함을 보였으며 제안하는 방법은 기존 TMIV 대비 PSNR 지표에서 평균 17.0%의 BD-rate 감소율을, IV-PSNR 지표에서는 14.6%의 BD-rate 감소율을 보였다. 현재 제안하는 방법은 사용자 시점에 대한 입력을 기반으로 하고 있지만, 향후 사용자 시선 예측 기술을 통해 효율적인 전송을 위한 연구를 계속 진행할 예정이다.

참고 문헌 (References)

- [1] ISO/IEC JTC1/SC29/WG11, "New Work Item Proposal on Coded Representation of Immersive Media," MPEG 117, N16541, Jan. 2017.
- [2] ISO/IEC JTC1/SC29/WG11/N18145, "Call for Proposals on 3DoF+ Visual", MPEG 125, Jan. 2019.
- [3] ISO/IEC JTC1/SC29/WG11/M41555, "Proposed architectures for supporting Windowed 6DoF, Omnidirectional 6DoF and 6DoF media", MPEG 120, Oct. 2017.

- [4] ISO/IEC JTC1/SC29/WG11/N19492, "Exploration Experiments on Processing for Future MPEG Immersive Video", MPEG 131, Jul. 2020.
- [5] ISO/IEC JTC1/SC29/WG11/N19485, "Description of MPEG Immersive Video Core Experiments 1", MPEG 131, Jul. 2020.
- [6] ISO/IEC JTC1/SC29/WG11/N19491, "Exploration Experiments on Coding for Future MPEG Immersive Video", MPEG 130, Apr. 2020.
- [7] ISO/IEC JTC1/SC29/WG11/M49590, Video-based Point Cloud Coding High Level Syntax: Updates and Unification with the Working Draft on Metadata for Immersive Video", MPEG 127, Jul. 2019.
- [8] ISO/IEC JTC1/SC29/WG11/M49406, "Group-Based TMIV", MPEG 127, Jul. 2019.
- [9] ISO/IEC JTC1/SC29/WG11/M54151, "Grouping and anchor study on MIV content", MPEG 131, Jun, 2020.
- [10] X. Corbillon, G. Simon, A. Devlic and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1-7. doi: <https://doi.org/10.1109/ICC.2017.7996611>.
- [11] J. Son, D. Jang, E-S. Ryu. "Implementing Motion-Constrained Tile and Viewport Extraction for VR Streaming." In Proceedings of the 28th ACM Network and Operating System Support for Digital Audio and Video (NOSSDAV'18), pp. 61-66, 2018. doi: <https://doi.org/10.1145/3210445.3210455>
- [12] S. Lee, D. Jang, J-B. Jeong, E-S. Ryu. "Motion-constrained tile set based 360-degree video streaming using saliency map prediction." In Proceedings of the 29th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'19), pp. 20-24, 2019. doi: <https://doi.org/10.1145/3304112.3325614>
- [13] ISO/IEC JTC1/SC29/WG11/M47749, "[V-PCC] A tile group design for V-PCC", A. M. Tourapis, J. Kim, K. Mammou, MPEG 126, Mar. 2019.
- [14] ISO/IEC JTC1/SC29/WG11/N18654, "Requirements for Immersive Media Access and Delivery", MPEG 127, Jul. 2019.
- [15] J. Jeong, S. Lee, I. Ryu, T. T. Le, and E. Ryu. "Towards Viewport-dependent 6DoF 360 Video Tiled Streaming for Virtual Reality Systems." In Proceedings of the 28th ACM International Conference on Multimedia (MM '20), pp. 3687 - 3695, 2020. doi: <https://doi.org/10.1145/3394171.3413712>
- [16] ISO/IEC JTC1/SC29/WG11/M54409, "[MPEG-I Visual] Ray-based blending weight for 6DoF view synthesis," MPEG 131, 2020.
- [17] ISO/IEC JTC1/SC29/WG11/N19483, "Test Model 6 for MPEG Immersive Video," MPEG 131, Jul. 2020.
- [18] ISO/IEC JTC1/SC29/WG11/N19484, "Common Test Conditions for MPEG Immersive Video," MPEG 131, Jul. 2020.
- [19] ISO/IEC JTC1/SC29/WG4/M57726, "AhG on MPEG Immersive Video", MPEG 136, Oct. 2021.
- [20] ISO/IEC JTC1/SC29/WG4/N0145, "Exploration Experiments on Future MPEG Immersive video", MPEG 136, Oct. 2021.
- [21] ISO/IEC JTC1/SC29/WG4/N0013, "Software manual of IV-PSNR for Immersive Video," MPEG 132, Oct. 2020.
- [22] J. Jeong, S. Lee, D. Jang and E. Ryu, "Towards 3DoF+ 360 Video Streaming System for Immersive Media," in IEEE Access, vol. 7, pp. 136399-136408, 2019. doi: <https://doi.org/10.1109/ACCESS.2019.2942771>.
- [23] S. Lee, J. -B. Jeong and E. -S. Ryu, "Efficient Group-Based Packing Strategy for 6DoF Immersive Video Streaming," 2022 International Conference on Information Networking (ICOIN), 2022, pp. 310-314. doi: <https://doi.org/10.1109/ICOIN53446.2022.9687139>.
- [24] Kwak, S., Yun, J., Jeong, J.-Y., Kim, Y., Ihm, I., Cheong, W.-S., and Seo, J., View synthesis with sparse light field for 6DoF immersive video, ETRI Journal 44 (2022), 24 - 37. doi: <https://doi.org/10.4218/etrij.2021-0205>
- [25] Jeong, J.-B., Lee, S. and Ryu, E.-S. (2021), Sub-bitstream packing based lightweight tiled streaming for 6 degree of freedom immersive video. Electron. Lett., 57: 973-976. doi: <https://doi.org/10.1049/ell2.12329>

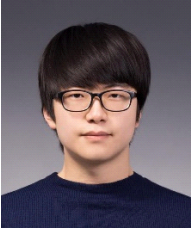
저 자 소 개



이 순 빈

- 2020년 3월 : 가천대학교 컴퓨터공학과 학사
- 2020년 3월 ~ 2022년 3월 : 성균관대학교 컴퓨터교육과 석사
- 2022년 3월 ~ 현재 : 성균관대학교 컴퓨터교육학과 박사과정
- ORCID : <https://orcid.org/0000-0002-8951-0335>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준

저 자 소 개



정 종 범

- 2018년 8월 : 가천대학교 컴퓨터공학과 학사
- 2018년 9월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 석박통합과정
- 2021년 8월 ~ 2022년 1월 : Purdue University 방문연구원
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



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 부교수
- ORCID : <https://orcid.org/0000-0003-4894-6105>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 코딩 및 국제 표준, HMD/VR 응용분야