

특집논문 (Special Paper)

방송공학회논문지 제28권 제2호, 2023년 3월 (JBE Vol.28, No.2, March 2023)

<https://doi.org/10.5909/JBE.2023.28.2.194>

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

## MPEG Immersive Video를 위한 그룹 기반 적응적 스트리밍

정 종 범<sup>a)</sup>, 이 순 빈<sup>a)</sup>, 최 재 열<sup>a)</sup>, 이 광 순<sup>b)</sup>, 곽 상 운<sup>b)</sup>, 정 원 식<sup>b)</sup>, 이 봉 호<sup>b)</sup>, 류 은 석<sup>a)‡</sup>

## Towards Group-based Adaptive Streaming for MPEG Immersive Video

Jong-Beom Jeong<sup>a)</sup>, Soonbin Lee<sup>a)</sup>, Jaeyeol Choi<sup>a)</sup>, Gwangsoon Lee<sup>b)</sup>, Sangwoon Kwak<sup>b)</sup>,  
Won-Sik Cheong<sup>b)</sup>, Bongho Lee<sup>b)</sup>, and Eun-Seok Ryu<sup>a)‡</sup>

### 요 약

다수의 색상 및 거리 순서쌍으로 구성된 몰입형 영상 압축을 위한 MPEG immersive video (MIV) 표준은 시점 간 중복 영역 제거 후 잔여 영상을 병합하여 높은 압축률을 확보하였다. 비슷한 영역을 표현하는 시점 간 그룹화를 통해 품질 향상 및 선택적 스트리밍 구현이 가능하나, 최근 그룹 기반 MIV 부호화 기술은 활발히 논의되고 있지 않다. 본 논문은 최신 MIV 참조 소프트웨어에서 그룹 기반 부호화 기술을 이식하고, 최적의 그룹 별 시점 및 영상 개수 산출을 위한 실험을 진행하였으며, 출력 영상 내 잔여 영상의 비율을 기반으로 전역적 영상 표현을 위한 최적의 출력 영상 수를 결정하는 기법을 제안한다.

### Abstract

The MPEG immersive video (MIV) coding standard achieved high compression efficiency by removing inter-view redundancy and merging the residuals of immersive video which consists of multiple texture (color) and geometry (depth) pairs. Grouping of views that represent similar spaces enables quality improvement and implementation of selective streaming, but this has not been actively discussed recently. This paper introduces an implementation of group-based encoding into the recent version of MIV reference software, provides experimental results on optimal views and videos per group, and proposes a decision method for optimal number of videos for global immersive video representation by using portion of residual videos.

Keyword : Metaverse, Virtual reality, MIV, 6DoF, Group-based encoding

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

b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)

‡ Corresponding Author : 류은석(Eun-Seok Ryu)

E-mail: [esryu@skku.edu](mailto:esryu@skku.edu)

Tel: +82-2-760-0677

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

※ This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by ICT R&D program of MSIT/IITP[2022-0-00022-001, Development of immersive video spatial computing technology for ultra-realistic metaverse services]

· Manuscript February 3, 2023; Revised March 3, 2023; Accepted March 3, 2023.

Copyright © 2023 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

몰입형 메타버스 구현이 주목받음에 따라, 더 현실적인 가상 및 증강 현실을 표현 가능한 부호화 및 전송, 렌더링 기술들이 주목받고 있다. 영상 압축 표준화 기관인 moving picture experts group (MPEG) 은 3차원 공간에서 점 단위로 정보를 표현하는 3-D 포인트 클라우드 (point cloud) 를 2차원 평면에 매핑 (mapping) 하여 종래에 널리 사용되던 2-D 영상 압축 표준 (e.g., high-efficiency video coding; HEVC) 을 사용하여 높은 압축률을 기록한 video-based point cloud coding (V-PCC) 표준화를 진행하였고, Nokia의 실시간 구현으로 V-PCC 상용화가 가능성이 증명되었다<sup>[1,2]</sup>. 또한 MPEG은 다수의 카메라를 통해 취득하여 다수의 객체 및 배경을 표현한 몰입형 영상에 대해 시점 간 압축을 진행하고, 잔여 정보를 2-D 영상 형태로 출력하여 V-PCC와 유사하게 종래의 2-D 영상 압축 표준을 이용해 호환성 및 높은 압축률을 확보한 MPEG immersive video (MIV) 표준화를 진행하였으며, ETRI의 실시간 구현으로 V-PCC와 마찬가지로 상용화가 가능성이 증명되었다<sup>[3-5]</sup>. MIV 부호기는 색상, 거리를 나타내는 텍스처 (texture) 및 지오메트리 (geometry) 영상의 쌍으로 구성된 다수의 몰입형 영상과 그에 대응하는 카메라 매개변수를 이용하여 각 시점을 기본 시점 (basic view) 또는 추가 시점 (additional view) 으로 분류한다. 이후 프루닝 (pruning) 을 통해 시점 간 중

복성이 제거되는데, 이 때 중복성은 오직 추가 시점에서만 제거된다. 기본 시점 및 추가 시점에서의 잔여 영상은 패킹 (packing) 을 통해 직사각형 형태의 패치 (patch) 로 추출되어 아틀라스 (atlas) 라 정의되는 출력 영상 내에 저장된다. 아틀라스는 HEVC 및 versatile video coding (VVC) 영상 압축 표준에 의해 부호화 및 복호화 될 수 있다<sup>[6,7]</sup>. 이후 MIV 복호기는 사용자가 시청하고 있는 가상 시점에 대해 시점 합성을 진행하고, 이를 통해 MIV는 사용자에게 머리 회전과 몸 움직임을 모두 지원하는 six degrees of freedom (6DoF) 를 지원할 수 있다.

그림 1은 MIV 부호기를 통한 아틀라스 생성 예시를 나타낸다. 그림에 도시된 테스트 시퀀스는 1-D 배열 형태로 이뤄진 30개의 카메라를 통해 농구장을 촬영한 CBAbasketball이고, 각 시점은 2048×1088의 해상도를 가진다<sup>[8]</sup>. 아틀라스 크기는 2048×4352로 설정되었고, 기본 시점 및 추가 시점 패치들은 패킹을 통해 아틀라스 내 빈 공간에 삽입되었다. 이 때 만약 삽입이 불가능할 경우 해당 패치는 2개의 작은 패치들로 분할되어 다시 패킹을 시도하며, 분할된 패치가 최소 패치 크기보다 작을 경우엔 해당 패치는 패킹되지 않고 버려진다. 아틀라스 내에 패킹된 패치들은 서로 다른 시점의 정보를 담고 있고, 시점 복원 및 가상 시점 합성을 위해서는 각 패치 정보들이 필요하다. MIV 부호기는 visual volumetric video coding (V3C) 표준에 따라 패치 및 아틀라스 정보들을 MIV 비트스트림 형태

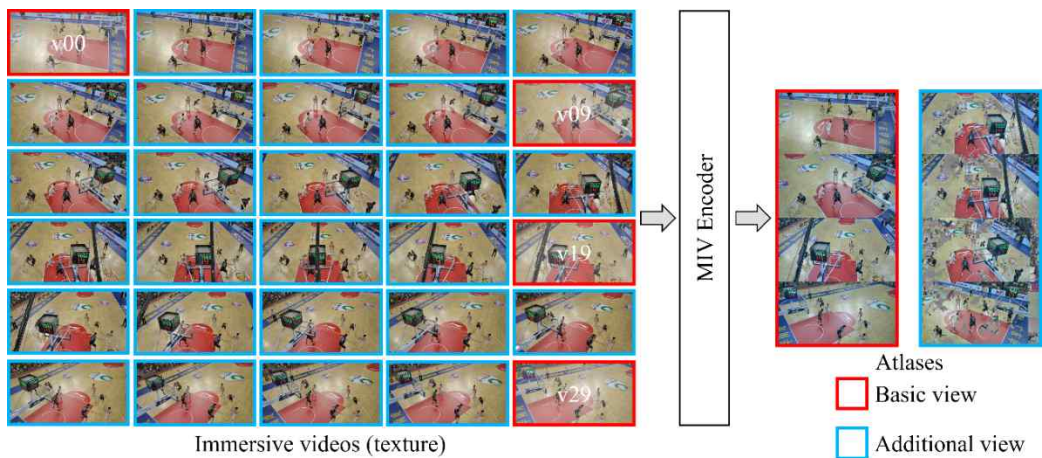


그림 1. 몰입형 영상을 이용한 아틀라스 생성 예시  
 Fig. 1. An example of atlas generation using immersive video

로 생성하여 저장하고, MIV 복호기는 이를 참조하여 시점 복원 및 가상 시점 합성을 진행한다<sup>9,10</sup>. MIV 부호화 효율 향상을 위해 지오메트리에 대한 비선형 양자화를 통해 부호화 효율을 향상시킨 연구<sup>11</sup>, 프레임 패킹 (frame packing) 을 통해 복호기 개수를 조절하는 연구<sup>12,13</sup>, 이중 프루닝을 통해 압축률을 개선한 연구<sup>14</sup>, ray 기반 렌더링 및 GPU 기반 뷰 합성을 통한 실시간화 연구<sup>15</sup> 등이 제안되었다.

그림 1에 도시된 바와 같이 MIV 부호기는 아틀라스 크기 및 개수가 주어지면 그에 맞춰 압축률을 최대화하기 위해 프루닝 및 패킹을 진행한다. 유사한 영역을 나타내는 시점들끼리 묶는 그룹화 이후 그룹별로 MIV 부호화를 진행하면 그룹 별로 기본 및 추가 시점이 결정되어 프루닝을 더 적게 하면서도 일관된 중복성 제거를 통해 렌더링 품질 향상을 이룰 수 있었고, 그룹 기반 MIV는 MIV 참조 소프트웨어인 test model for immersive video (TMIV) 에서 구현 및 채택되었다<sup>16</sup>. 그룹 기반 MIV의 효율은 검증되었으나<sup>17,18</sup>, 그룹 분류 알고리즘, 그룹 별 부호화 및 메타데이터 생성, 그룹 기반 복호화 기능을 TMIV에서 구현하고 최신 버전을 유지하는 과정의 어려움으로 최신 버전의 TMIV에서는 그룹 기반 부호화 기술이 제외되었으며 활발히 논의되고 있지 않다. 그러나, 그룹 기반 MIV 부호화 기술은 그림 1에 도시된, 대형 공연장에서의 실사 콘텐츠 처리 시 압축률 향상 및 복호화 복잡도 감소를 위해 사용될 수 있으며, 최신 버전의 TMIV에서 논의될 필요가 있다.

본 논문은 MIV의 그룹 기반 부호화 및 복호화 기술을 최신 버전의 TMIV에서 구현하고, 대형 공간 콘텐츠에 대

한 효율을 검증한다. 그림 2는 그룹 기반 MIV 부호기의 동작을 도시한다. 텍스처와 지오메트리 쌍으로 구성된 몰입형 영상은 카메라 매개변수와 함께 Group Selector에 입력되고, 이후 영상 위치 및 시야각에 기초하여 다수의 그룹으로 분할된다. View Optimizer는 몰입형 영상을 기본 및 추가 시점으로 분류하고, Atlas Constructor는 프루닝 및 패킹 진행 후 아틀라스를 생성하고 메타데이터를 MIV 비트스트림 형태로 저장한다. Group Metadata Merger는 그룹 단위로 출력되는 메타데이터를 묶어 단일 MIV 비트스트림으로 저장하고, Video Encoder는 아틀라스를 부호화한 후 상기 MIV 비트스트림과 멀티플렉싱 (multiplexing) 하여 단일 비트스트림을 생성한다. 또한 본 논문은 그룹 기반 MIV 부호화 시 반드시 고려되어야 할 그룹 분할 수 및 그룹 당 아틀라스 수에 따른 성능 분석을 소개하고, 아틀라스 내 패치 점유율 및 아틀라스 내에 포함되지 않아 버려지는 패치 비율을 로그로 출력하는 기능을 구현하여 향후 아틀라스 및 기본 시점 수를 결정 시 참고할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 배경 연구로 그룹 기반 MIV 부호화 기법을 소개한다. 3 절에서는 최신 버전 TMIV에서의 그룹 기반 부호기/복호기 및 로그 출력 구현 내용을, 4 절에서는 실험 결과를, 마지막으로 5 절에서는 본 논문의 결론을 서술한다.

## II. 배경 연구

2019년에 TMIV v1.0이 발표된 후<sup>19</sup>, MIV 표준화 그룹

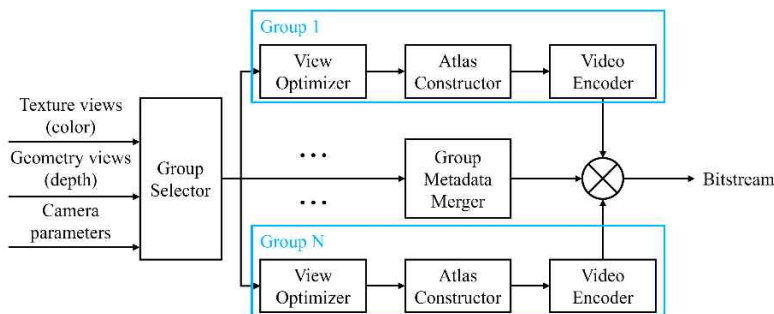


그림 2. 그룹 기반 MIV 부호기 블록 다이어그램  
Fig. 2. Block diagram of group based MIV encoder

은 MIV의 성능 향상을 위한 core experiment (CE) 및 exploration experiment (EE) 를 진행해왔다. 2019년에 MIV 표준화 그룹이 진행한 CE1은 적응적 입력 영상/가상 시점 영상 선택 및 처리를 목적으로 진행되었고<sup>[20]</sup>, Intel에 의해 그룹 기반 MIV가 제안되었다<sup>[16]</sup>. 그림 1에 도시된 MIV 부호화 과정의 경우 아틀라스 개수가 2개로 제한되었고, 아틀라스 해상도는 2048×4352인데 반해 입력으로 주어진 몰입형 영상은 2048×1088 해상도의 영상이 30개가 존재하므로 프루닝이 이루어져도 여전히 아틀라스 크기 대비 많은 양의 픽셀들이 잔류하여 모든 패치들이 아틀라스 내에 패키징될 수 없는 문제가 발생한다. 이 경우 MIV 부호기는 압축률을 최대화하기 위해 프루닝을 다시 진행하여 잔여 영상 픽셀 수를 낮추거나, 패치를 분할하여 패키징을 시도하다가 정의된 최소 크기보다 작게 분할될 경우 패치를 누락시켜 아틀라스 내에 최대한 많은 양의 패치를 삽입하게 된다. 이 경우 MIV 부호기가 아틀라스를 이용해 가상 시점 영상을 합성 시 배경을 포함한 패치가 전경을 표현하는 물체 위에 렌더링되어 합성된 시점의 품질이 하락하는 문제가 발생하였다. Intel이 제안한 그룹 기반 MIV는 그룹 단위로 기본

및 추가 시점이 분류되고, 이후 프루닝 및 패키징이 진행되어 일관되게 지역적 특성을 보존하고 합성된 시점의 품질을 향상시킬 수 있었다. 그룹 기반 MIV는 TMIV v3.0에 채택되었고, CE가 진행됨에 따라 프루닝 및 가상 시점 합성 시의 그룹 적용 효과에 대한 실험이 진행되었다<sup>[21]</sup>. 그림 3은 그룹 기반 TMIV 부호기의 상위 다이어그램을 도시한다. 입력으로 주어진 몰입형 영상들은 그룹 단위로 분할되어 부호화가 그룹별로 진행되고, 텍스처, 지오메트리, 그리고 아틀라스 내 유효 영역을 나타내는 점유 지도 (occupancy map) 가 영상 부호기에 입력되어 부호화 될 수 있다. 이렇듯 TMIV는 그룹 단위 부호화 개념을 채택하고 있으나, MIV 표준화가 진행되면서 소스 코드 관리의 어려움으로 인해 TMIV v9.0 이후로는 그룹 기반 MIV가 TMIV 내에서 제거되었다.

합성된 시점 품질 향상 외에도, 그룹 기반 MIV는 선택적 스트리밍 시 대역폭을 크게 절감할 수 있다. 종래의 2-D 영상 대비 해상도가 크고 head-mounted display (HMD) 를 통해 표시되어 고화질을 요구하는 360도 영상 전송 시 사용자 시점은 고화질로, 그 외 영역은 저화질로 전송하는 사용

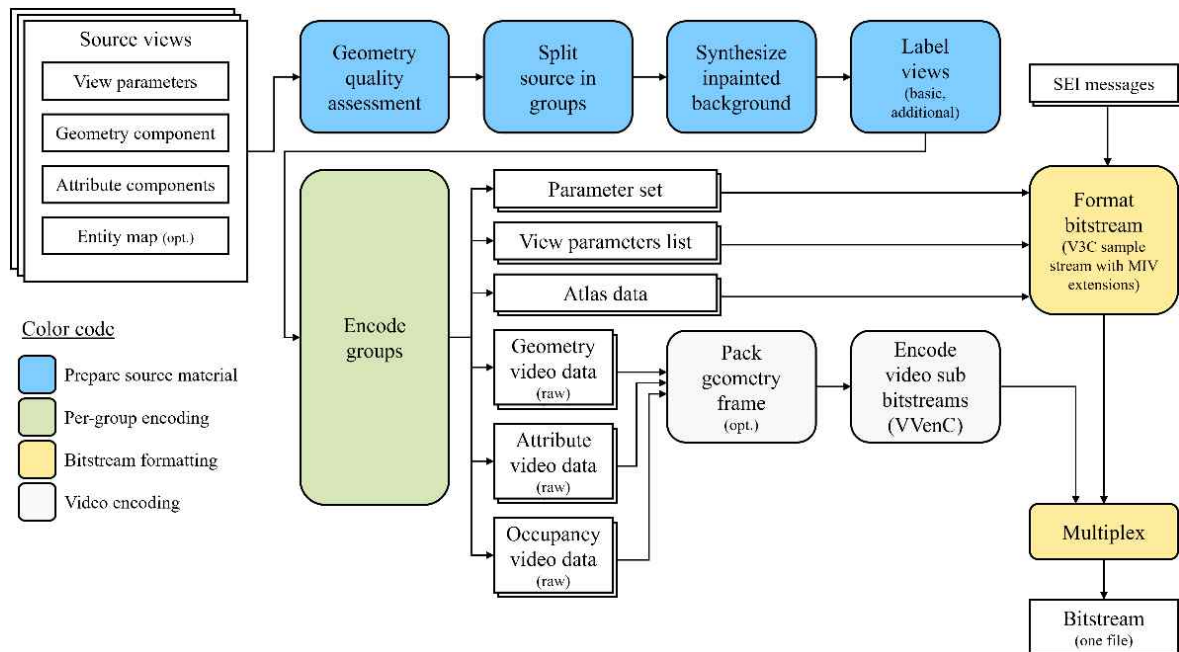


그림 3. 그룹 기반 TMIV 부호기 상위 다이어그램  
 Fig. 3. High-level block diagram of group based TMIV encoder

자 시점 기반 스트리밍 연구가 제안되었다<sup>[22,23]</sup>. 통상의 단일 360도 영상보다 더 높은 대역폭을 요구하는 몰입형 영상 전송 시에도 사용자 시점 기반 전송 사용은 유용하게 사용될 수 있다. 그룹 기반 MIV를 사용할 경우, 인접한 시점을 그룹으로 묶어 그룹 단위로 프루닝 및 패키징이 진행되므로 그룹 별 복원이 가능하며, 따라서 사용자가 시청하는 영역을 포함하는 그룹만 전송이 가능하다. 특히 대규모 카메라 배열을 통해 취득된 몰입형 영상의 경우 시점 간 유사도가 크지 않고 그룹을 미적용 시 모든 아틀라스에 접근하여 시점 합성을 진행할 경우 복잡도가 높으므로 그룹 기반 MIV

가 유용하게 사용될 수 있다.

### III. MIV 그룹 기반 스트리밍

본 절은 최신 버전 TMIV에서의 그룹 기반 부호기/복호기 구현 내용을 기술하고, 대형 공간 카메라 배열을 통해 취득된 몰입형 영상에 대한 아틀라스 수 결정을 위한 로그 출력 기법에 대해 설명한다. 그림 4는 그룹 적용/미적용에 따른 몰입형 영상 프루닝 그래프 및 아틀라스 예시를 도시

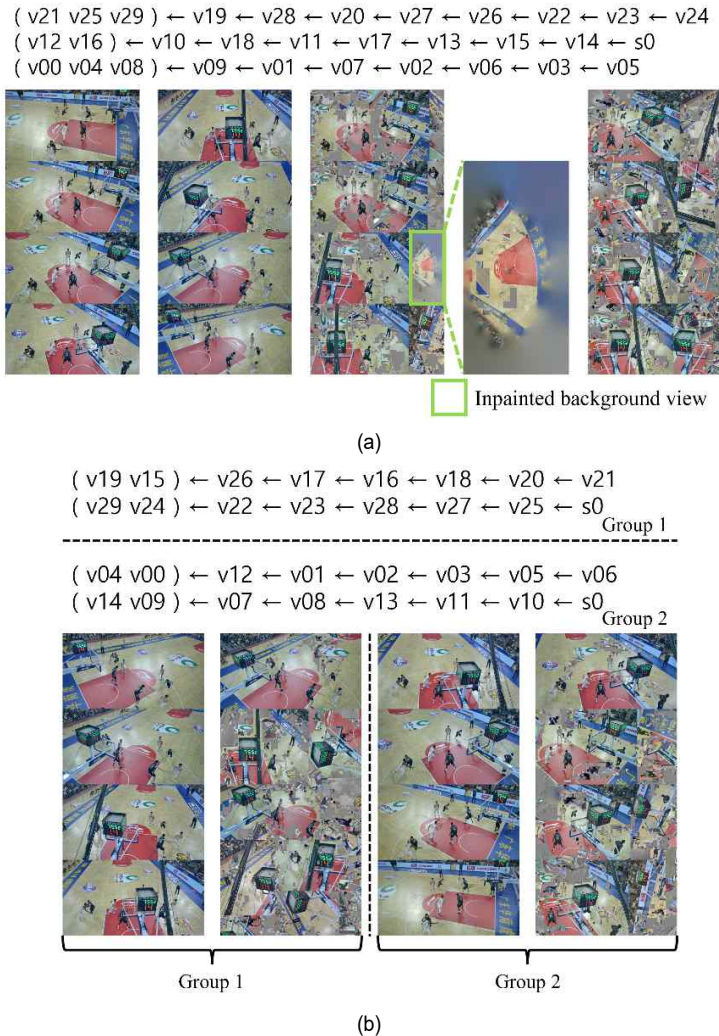


그림 4. 그룹 적용/미적용 시 프루닝 그래프 예시. (a) 그룹 미적용 시 프루닝 그래프 예시, (b) 그룹 적용 시 프루닝 그래프 예시  
 Fig. 4. Examples of pruning graph with/without group. (a) pruning graph without group, (b) pruning graph with group

한다. 테스트 시퀀스로 CBAbasketball이 사용되었고, 아틀라스 개수는 4개로 설정되었다. 프루닝 그래프는 시점 간 프루닝 순서를 나타내며, 괄호로 묶인 시점들은 기본 시점을, 그 외 시점들은 추가 시점으로 표현되었다. 그림4(a)는 그룹 미적용 시 생성된 3개의 프루닝 그래프를 도시한다. MIV 부호기는 시점들을 클러스터 (cluster) 단위로 분할하여 클러스터 단위로 프루닝을 진행하는데, 이를 통해 프루닝 시 기본 시점의 개수가 많아지는 것을 방지하고 프루닝 그래프를 단순화하여 복잡도를 절감한다. 프루닝 그래프에 도시된 좌측 시점부터 우측 시점 순으로 프루닝이 진행되고, 추가 시점 간 프루닝도 진행된다. 그림 4(a)의 첫 번째 클러스터의 경우 시점 v19를 프루닝할 때 v21, v25, v29만 사용되지만, 시점 v24를 프루닝할 때 v21, v25, v29, v19, v28, v20, v27, v26, v22, v23 모두가 사용된다. 그림 4(a)의 두 번째 클러스터에는 시점 s0이 존재하는데, 아틀라스 내에 초록색으로 강조되었다. 이는 입력으로 주어진 몰입형 영상들을 이용하여 인페인팅 (inpainting) 된 배경 시점을 합성하여 배경 정보를 전역적으로 표현하고 합성된 시점의 품질을 향상시키기 위해 제안되었고, TMIV에 반영되었다<sup>[24]</sup>. 그림 4(b)는 그룹 적용 시 생성된 프루닝 그래프 및 아틀라스를 도시한다. 예시엔 그룹 2개가 사용되었고, 그룹 하나 당 두 개의 아틀라스가 할당되었다. 각 그룹은 점선으로 구분되며, 프루닝 및 패킹은 그룹별로 진행된다.

MIV 표준화 그룹은 common test conditions (CTC) 를 통해 실험에 사용되는 소프트웨어의 버전, 입력 몰입형 영상인 테스트 시퀀스, 아틀라스 개수 등을 정의한다<sup>[25]</sup>. MIV CTC는 2개의 4K 텍스처 아틀라스와 그에 대응하는 지오메트리 아틀라스를 사용할 것을 권고하는데, 이는 추후 MIV를 모바일 디바이스에서 구현 시에도 충분한 성능을 보장하기 위해서임이 141차 MPEG 회의에서 언급된 바 있다. 그러나, 그림 4에서 예시로 도시된 CBAbasketball의 경우 시점 당 해상도가 2048×1088이고, 시점 수가 30개에 달하여 2개의 아틀라스 내에 모든 패치들이 패킹되지 않아 영상 부호화 시 고품질로 부호화 하여도 복호화 및 시점 합성 시 홀 (hole) 이 발생하는 문제가 있다. 따라서 몰입형 영상의 특성을 고려한 아틀라스 개수 결정 기법이 필요하며, 그룹 수 및 그룹 별 아틀라스 수도 고려되어야 한다. 본 절의 구성은 다음과 같다. 3.1 절에서는 최신 버전 TMIV에서의 그룹 기반 부호기/복호기 구현 내용을 설명한다. 3.2 절에서는 그룹 기반 부호화 시 아틀라스 수 결정을 위한 아틀라스 내 패치 점유율 및 패킹되지 않고 버려지는 패치 비율을 출력하는 기능 구현에 대해 상세히 기술한다.

### 1. 그룹 기반 부호기/복호기 구현

본 절은 최신 버전의 TMIV에서의 그룹 기반 부호기/복

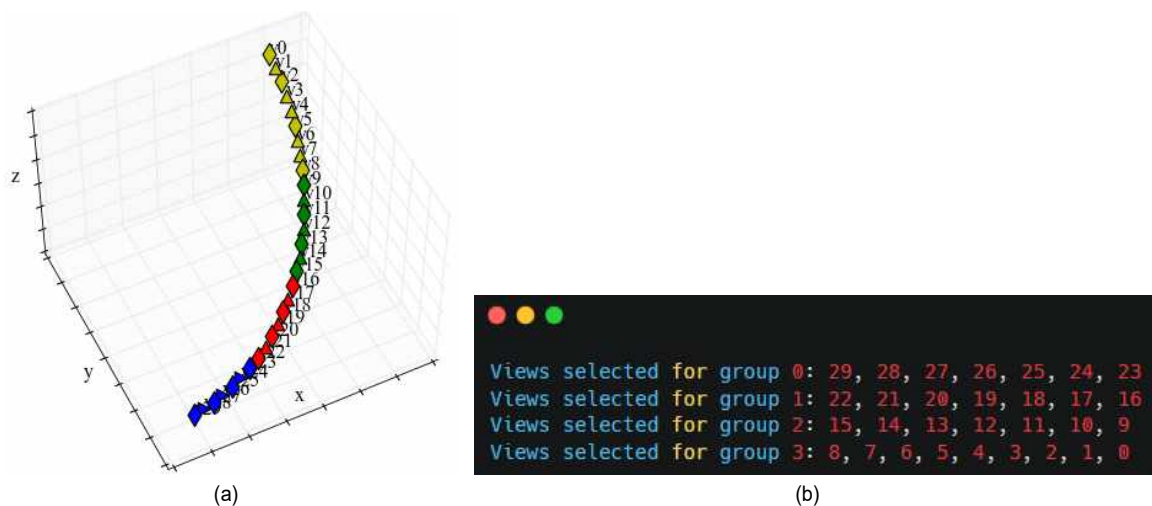


그림 5. 그룹 별 시점 분할 예시. (a) 시점 좌표 시각화, (b) 그룹 별 시점 출력 예시  
 Fig. 5. Example of views splitting per group. (a) visualization of views, (b) views for each group

호기 구현에 대해 설명한다. 2019년 5월에 TMIV v1.0이 제안된 이후, 2022년 7월에 TMIV v14.0이 제안되었다. 본 논문은 TMIV v13.1에서 그룹 기반 부호기 및 복호기를 구현하였다. 2021년 10월에 TMIV v11.0이 나온 이래로 TMIV v13.0에서 부호화/복호화 함수 호출 방식, V3C 표준 메타데이터를 표현하는 구조체 표현 방식 등의 구현 내용에서 상당한 변화가 있었고, 이후 구조체, 클래스, 함수 호출 등의 방식은 큰 차이가 없었으므로 안정성이 보장된 v13.1을 사용하였다.

표 1. ISO/IEC 23090-12 그룹 매핑 선택스  
Table 1. group mapping syntax of ISO/IEC 23090-12

group_mapping() {	Descriptor
gm_group_count	u(4)
if ( gm_group_count > 0 )	
for ( k = 0; k <= vps_atlas_count_minus1; k++ )	
j = vps_atlas_id[ k ]	
gm_group_id[ j ]	u(v)
}	

그림 3에 도시된 바와 같이, 그룹 기반 TMIV 부호기는 지오메트리 품질을 측정 후 소스를 그룹 별로 나눈다. 몰입형 영상의 x, y, z 좌표가 입력되면, TMIV 부호기는 입력된 몰입형 영상의 3차원 축들 중 제일 넓은 값을 가지는 축을 중심 축으로 정의한다. 제일 큰 중심 축을 가지는 시점이 키 포지션 (key position) 으로 정의되고, 모든 시점들과 키 포지션 사이의 거리값을 구한 후 상기 거리값에 따라 시점을 오름차순으로 정렬한다. 그룹 별 시점 수는 전체 시점 수를 그룹 수로 나누고 내림한 값으로 정의되고, 정렬된 시점의 집합에서 제일 거리값이 작은 순서대로 현재 그룹에 시점을 포함하며 그룹 별 시점 수만큼 이를 반복한다. 그룹 별로 시점을 할당하는 기능은 TMIV 부호기 단에서 sourceSplitter() 함수를 이용해 구현되었다. 그림 5는 sourceSplitter() 함수를 통한 그룹 별 시점 분할 예시를 나타낸다. CBAbasketball 테스트 시퀀스에 대해 4개의 그룹으로 분할을 진행하였고, 그림 5(a)에 도시된 바와 같이 3차원 공간에서 유사한 위치의 시점들끼리 그룹화가 되었다. 그림 5(b)는 sourceSplitter() 함수 실행 시 각 그룹 별 시점 인덱스를 출력한 결과를 나타낸다. 시점 v29가 키 포지션으로 정의되었고, 상기 그룹 별 시점 수 계산법에 따라 마지막 그룹을 제외한 그룹 별 시점 수는 7개이므로 첫 번째 그룹에는 시점 v29부터 시점 v23까지 포함되었다. 이후 마지막 그룹에는 나머지 9개의 시점들이 포함되었다.

모든 그룹에 대해 시점이 할당되면, 각 그룹별로 프루닝 및 패키징이 진행되고, 이후 TMIV 부호기는 그룹별로 생성된 V3C 메타데이터를 병합하여 통합된 MIV 비트스트림을 출력한다. 그룹별 V3C 메타데이터를 병합한다는 것은 그룹 수 등의 정보가 별도로 시그널링 되어야 함을 의미한다. ISO/IEC 23090-5 표준에 의해 정의된 V3C 표준의 확장으로 MIV는 ISO/IEC 23090-12 표준에 의해 선택스를 정의하는데, 상기 표준의 8.3.2.2 절은 그룹 개수 및 각 아틀라스가 어느 그룹에 속하는지를 명시하여 그룹 정보를 표현한다. 표 1은 ISO/IEC 23090-12 그룹 매핑 선택스를 나타낸다. gm\_group\_count는 그룹의 수를 나타내고, 부호 없는 4비트 정수로 표현되므로 0~15 사이의 값을 표현할 수 있다. gm\_group\_count가 0보다 클 경우, TMIV 부호기는 아틀라스 수 만큼 반복문을 수행하고, 변수 j에 아틀라스의 ID를 대입한다. 이후 gm\_group\_id[ j ]에는 아틀라스 j가 해당하는 그룹의 인덱스 값이 설정된다. 그림 5에서는 gm\_group\_count가 4이고, gm\_group\_id[ 0 ], gm\_group\_id[ 1 ], gm\_group\_id[ 2 ], gm\_group\_id[ 3 ]에는 각각 0, 1, 2, 3의 값이 대입된다.

ISO/IEC 23090-12에 따르면, TMIV 복호기는 복호화된 아틀라스 및 MIV 비트스트림을 입력받고 그룹 추출 과정

표 2. ISO/IEC 23090-12 그룹 추출 과정  
Table 2. group extraction process of ISO/IEC 23090-12

for ( a = 0; a < vps_atlas_count_minus1 + 1; a++ ) {
if ( AtlasGroupId[ a ] == targetGroupID )
invoke subclause 9.8.2 of ISO/IEC 23090-5 with targetAtlasId equal to a
}

을 거친다. 표 2는 ISO/IEC 23090-12 에서의 그룹 추출 과정을 나타낸다. targetGroupID는 추출하고자 하는 그룹의 ID를, AtlasGroupID[ a ] 는 현재 탐색중인 그룹의 ID를 나타낸다. 만약 두 값이 같을 경우, TMIV 복호기는 ISO/IEC 23090-5의 9.8.2 절에 명시된 작업인 V3C 유닛 추출을 진행한다. 본 논문은 그룹 기반 부호화 및 복호화 시 그룹 미 적용 대비 효율 및 그룹 당 아틀라스 개수에 따른 효율을 검증하였기에, 특정 그룹을 추출하는 실험은 진행하지 않았으나 ISO/IEC 23090-12가 제안하는 실시예로 구현이 가능하다.

## 2. 아틀라스 내 패치 점유율 및 누락된 패치 비율 출력 기능 구현

본 절은 최신 버전의 TMIV에서의 아틀라스 내 패치 점유율 및 누락된 패치 비율 출력 기능 구현에 대해 설명한다. 2022년 7월에 한국전자통신연구원과 한양대학교 공동연구팀은 아틀라스 크롭핑 (cropping) 에 대한 연구를 제안하였다<sup>[26]</sup>. 그림 6은 CBAbasketball 테스트 시퀀스에 대해 MIV 부호화 시 아틀라스 6개를 사용했을 때의 아틀라스 크롭핑 예시를 도시한다. 그림 6(a)에 도시된 바와 같이, 마지막 아틀라스에는 패치가 점유하지 않은 빈 공간이 존재한다. 빈 공간이 아틀라스에 포함되면 부호화 시 비트율의 소폭 증가 및 복호화 시 불필요한 픽셀 레이트 (pixel rate) 증가로

인한 복호화 시간 소폭 증가를 야기한다. 따라서, 제안된 연구는 패치가 점유한 영역을 검출하고 상기 영역만 크롭핑하여 오직 유효 영역만을 가지는 아틀라스들을 생성한다. 또한 상기 크롭핑은 프레임별로 픽처 크기가 변하는 것을 방지하고 부호기 및 복호기에서의 일관성을 유지하기 위해 group of pictures (GOP) 단위로 이루어진다. 제안된 연구는 24개의 시점을 가지고, 각 시점의 해상도가 2048×2048인 Museum 시퀀스에 대해 실험을 진행하였다. 상기 방법 적용 후 MIV 부호화 시 1.6%의 시간 증가가 있었으나, 영상 압축 도구를 이용한 아틀라스 부호화 시 3.0%의 시간 감소 및 아틀라스 복호화 및 가상 시점 영상 합성 시 9.9%의 시간 감소가 관찰되었고, TMIV에 최종 반영되었다.

본 논문은 [26]이 아틀라스 내의 패치 점유율을 계산하고 로그로 출력하는 점에 착안하여, 아틀라스 수를 결정 시 상기 정보를 참고할 수 있도록 하였다. CBAbasketball과 같이 시점 수가 많고 개별 시점이 고해상도인 테스트 시퀀스는 MIV CTC가 권장하는 2개의 아틀라스를 사용하여 MIV 부호화 시 모든 패치들이 아틀라스 내에 패키징될 수 없어 일부 패치들은 버려지는 현상이 발생한다. 이는 아틀라스 부호화 시 품질과 상관없이 홀 발생을 야기하여 객관적, 주관적 화질을 하락시키는 원인이 된다. 따라서, 아틀라스 내의 패치 점유율이 100%에 수렴한다면 기존에 사용했던 아틀라스 개수보다 더 많은 수의 아틀라스를 사용하여 MIV 부호화를 진행 시 홀 발생을 방지할 수 있다. 본 논문은 [26]이

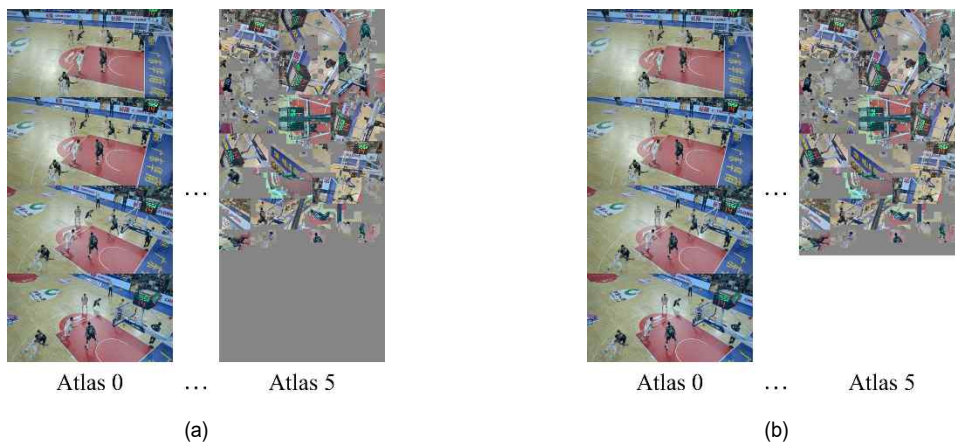


그림 6. 아틀라스 크롭핑 예시. (a) 크롭핑 전 아틀라스, (b) 크롭핑 후 아틀라스  
 Fig. 6. Example of atlas cropping. (a) atlases before cropping, (b) atlases after cropping



제안하지 않았던, 버려지는 패치들의 면적을 계산하여 로 그로 출력하는 기능을 구현하였고, 이를 통해 몰입형 영상 별로 홀이 발생하지 않게 하기 위한 적정 아틀라스 수를 탐색 시 참고할 수 있도록 하였다.

MIV 부호기는 패치를 아틀라스 내에 패킹 시 다음과 같은 과정을 거친다. MIV 부호기는 프루닝 이후 직사각형 형태의 패치들을 검출하여 정렬한다. 정렬 방식은 면적에 따른 내림차순 정렬과 시점 ID에 따른 오름차순 정렬이 있고, MIV CTC는 전자를 권장한다. TMIV 내에서는 패치들을 우선순위 큐 (priority queue) 내에 저장하는 방식으로 정렬을 수행하였다. 이후 MIV 부호기는 2가지 패킹 방법을 순차적으로 시도한다. 먼저 MIV 부호기는 모든 아틀라스 내에 이미 다른 패치가 패킹된 공간 내에 새 패치가 패킹될 수 있는지 확인하고, 가능하면 해당 위치에 새 패치를 삽입한다. 패치는 직사각형 형태로 추출되지만, 실제로 패치 내에 표현된 잔여 영상들은 직사각형 형태가 아니기에 빈 공간이 존재할 수 있으며, 여기에 작은 패치가 삽입되면 패치 내 유효 영역을 침범하지 않으면서 아틀라스 내 공간을 좀 더 효율적으로 사용 가능하다. 패치 내에 패치를 삽입하는 기능은 TMIV v1.0부터 지원되었고, 최신 버전의 TMIV에서는 enablePatchInPatch 파라미터의 값을 true로 설정하여 사용 가능하며 MIV CTC도 상기 값을 true로 설정할 것을 권장한다. 그림 7은 패치 패킹 기법들의 예시를 나타낸다. 그림 7(a)는 패치 내에 패치를 삽입 시 생성된 아틀라스를

도시한다. 빨강 점선으로 표시된 영역을 보면 큰 패치 내에 회색으로 표현된 빈 공간에 작은 패치들이 삽입된 것을 볼 수 있다. 만약 기존 패치 내에 새로운 패치 삽입이 불가능하면 MIV 부호기는 새로운 패치를 90도 회전하여 패킹될 수 있는지 확인한다. 패치를 회전시켜도 삽입이 불가능하면, MIV 부호기는 아틀라스 내 빈 공간에 패치를 삽입할 수 있는지 확인하고, 삽입이 불가능하면 패치를 90도 회전하여 패킹을 시도한다. 그림 7(b)는 패치 내 패치 패킹 기능을 비활성화 시 생성된 아틀라스를 도시한다. 아틀라스 내 패치들은 서로 다른 공간에 표현되고, 패치 내 패치 패킹을 적용했을 때 대비 아틀라스 내 공간을 많이 사용해야 함을 알 수 있다. 상기 방법들을 모두 시도했을 때 패킹이 불가능하면, MIV 부호기는 패치를 분할하여 2개의 작은 패치들을 생성한다. 이후 분할된 패치들이 미리 설정된 패치 최소 면적보다 크면 우선순위 큐에 삽입하여 다시 패킹을 진행하고, 그렇지 않을 경우 삽입을 하지 않아 해당 패치는 버려진다. 본 논문은 패치 삽입이 실패하여 상기 패치가 2개의 작은 패치들로 분할되었을 때, 2개의 패치들을 각각 검사하여 최소 크기보다 작거나 같을 경우 상기 패치의 면적을 기록하는 기능을 구현하였다. 만약 몰입형 영상의 해상도와 시점 수 대비 적은 수의 아틀라스를 사용하여 MIV 부호화를 진행하였을 경우, 아틀라스 내 패치 점유율 및 버려진 패치 면적 로그들을 참고하여 패치 누락이 발생하지 않도록 아틀라스 수를 조정할 수 있다.

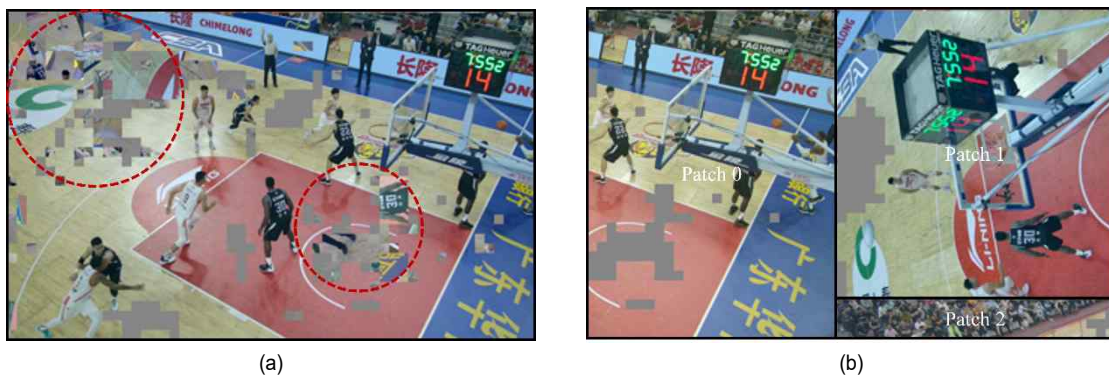


그림 7. 패치 패킹 예시, 검정 실선은 단일 패치를, 빨강 점선은 패치 내에 패킹된 패치들의 집합을 나타냄. (a) 패치 내 패치 패킹 적용 예시, (b) 패치 내 패치 패킹 미적용 예시

Fig 7. Example of patch packing, a box with black line represents a single patch, and a red-dotted circle represents a set of patches in a patch. (a) example when enabled patch in patch, (b) example when disabled patch in patch

#### IV. 실험 결과

본 절은 제안하는 그룹 기반 MIV의 효율과 패치 점유율 및 버려진 패치 비율을 기반으로 그룹 별 최적의 아틀라스 개수 결정 과정을 설명한다. 3.1 절에서 설명되었듯 그룹 기반 MIV 부호기/복호기는 TMIV v13.1 버전에서 구현되었다. MIV 부호기는 몰입형 영상을 입력 받아 MIV CTC가 권장하는 17 프레임에 대해 아틀라스 및 MIV 비트스트림을 생성한다. 이후 아틀라스는 HEVC 참조 소프트웨어인 HEVC test model (HM) 16.16 버전의 부호기에 의해 부호화되었다. 부호화 시 텍스처 아틀라스에 대해 22, 27, 32, 37, 42의 값을 가지는 양자화 매개변수 (quantization parameter; QP) 를 사용하였고, 지오메트리 아틀라스에 대해서는 MIV CTC에서 권장하는 텍스처 - 지오메트리 간 QP 매핑을 수행하는 수식을 통해 3, 7, 11, 15, 19의 QP 값을 사용하였다. HM 복호기가 비트스트림을 입력 받아 복호화된 아틀라스를 생성하였고, MIV 복호기는 상기 복호화된 아틀라스 및 MIV 비트스트림을 입력 받아 몰입형 영상의 각 시점에서의 가상 시점 영상 합성을 진행하였다. 원본 영상과 합성된 영상 간 품질 평가를 위해 peak signal-to-noise ratio (PSNR) 및 몰입형 영상 품질 측정을 위해 Poznan University of Technology (PUT) 연구팀에서 개발한 immersive video PSNR (IV-PSNR) 이 사용되었다<sup>[27]</sup>.

테스트 시퀀스로는 CBAbasketball이 사용되었다. 그림 8은 본 실험에 사용된 테스트 시퀀스의 카메라 배열 및 시점을 도시한다. 그림 8에 도시된 바와 같이, CBAbasketball은

표 3. 실험 조건

Table 3. Experimental Conditions

Condition	No. of atlases	No. of groups	Comparison group
Anchor_atlas2	2	-	-
Anchor_atlas4	4	-	-
Anchor_atlas6	6	-	-
Anchor_atlas8	8	-	-
Group2_atlas2	4	2	Anchor_atlas4
Group4_atlas1	4	4	Anchor_atlas4
Group3_atlas2	6	3	Anchor_atlas6
Group6_atlas1	6	6	Anchor_atlas6
Group4_atlas2	8	4	Anchor_atlas8

2048×1088 해상도를 가지는, 수렴형 호 형태로 1m 간격 단위로 배치된 30개의 2-D 카메라를 사용하여 농구 경기를 실사 형태로 취득하였다. 표 3은 본 실험의 조건들을 나타낸다. 실험에 사용된 두 시퀀스 모두 대형 공간 또는 다수의 카메라를 사용하여 MIV CTC가 권장하는 2개의 아틀라스를 사용 시 모든 패치가 패킹될 공간이 부족해 패치 누락이 발생한다. 따라서 본 실험에서는 2, 4, 6, 8개의 아틀라스를 사용하였고, 그룹을 적용하지 않은 실험 조건들을 대조군 (anchor) 으로 설정하였다. ‘Anchor\_atlas2’라는 실험 조건은 그룹을 적용하지 않고 아틀라스 2개를 사용한 실험 조건을 나타낸다. 그룹을 적용 시에는 그룹 수와 그룹 별 아틀라스 수도 조절할 수 있고, 본 실험에서는 총 5가지의 그룹 기반 실험 조건을 설정하였다. ‘Group2\_atlas2’라는 실험 조건은 입력된 몰입형 영상을 2개의 그룹으로 분할하고, 그룹 당 아틀라스가 2개 할당되어 총 아틀라스 수는 4개임을 의미한다. MIV 부호기는 maxBasicViewFraction 옵션

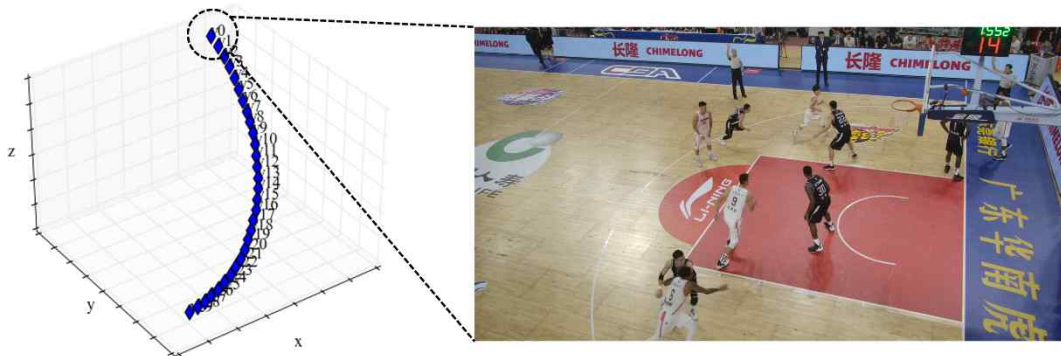


그림 8. 카메라 및 시점 v0, 시퀀스 CBAbasketball  
 Fig. 8. Camera coordination and view v0 of CBAbasketball

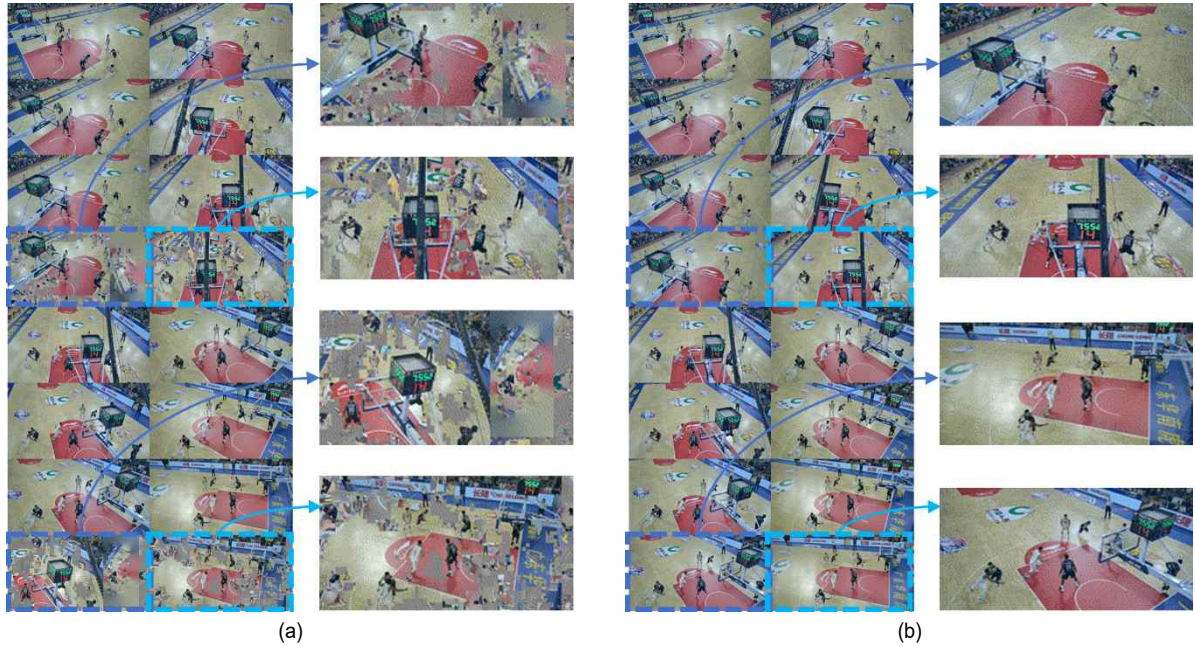


그림 9. 그룹 4개 분할, 그룹 당 아틀라스 2개 할당 시 기본 시점 수에 따른 아틀라스 예시. (a)  $\text{maxBasicViewFraction}=0.4$ 일 때의 아틀라스 예시, (b)  $\text{maxBasicViewFraction}=0.5$ 일 때의 아틀라스 예시  
 Fig. 9. Examples of atlases when using four groups and two atlases for each group. (a) atlases when  $\text{maxBasicViewFraction}$  is 0.4, (b) atlases when  $\text{maxBasicViewFraction}$  is 0.5

션을 통해 기본 시점의 개수를 조절하고, MIV CTC는 상기 옵션에 0.5의 값을 권장하므로 통상 기본 시점은 아틀라스 내 50%의 공간을 차지한다 ‘Anchor\_atlas8’ 및 ‘Group4\_atlas2’ 실험 조건은 아틀라스를 8개 사용하고, CBAbasketball 시퀀스의 기본 시점은 MIV CTC에서 권장하는 아틀라스 1개에 최대 4개가 포함되므로 총 4개의 아틀라스에 16개의 기본 시점이 포함되어야 한다. 그러나 ‘Anchor\_atlas8’ 실험 조건에서  $\text{maxBasicViewFraction}$  옵션에 0.5를 할당 시 기본 시점을 패킹하는 과정에서 MIV 부호기가 중지되는 문제가 발생하여, 본 실험에서는 아틀라스를 8개 사용하는 실험 조건에 한해 상기 옵션에 0.4를 할당하였다. 그림 9는 ‘Group4\_atlas2’ 실험 조건에 대한  $\text{maxBasicViewFraction}$  옵션에 따른 아틀라스 예시를 나타낸다. MIV 부호기는 기본 시점을 먼저 아틀라스에 삽입하기 때문에 통상 기본 시점들은 첫 번째 아틀라스에 포함된다. 따라서 그림 9는 아틀라스 내의 기본 시점들을 나타내기 위해 각 그룹에서의 첫 번째 아틀라스만 도시하였다. 그림 9(a)는 상기 옵션에 0.4의 값을 할당 시 아틀라스를 도시

한다. 총 12개의 기본 시점이 포함되었고, 각 그룹별로 첫 번째 아틀라스에 3개의 기본 시점이 포함되었고 아틀라스 하단에는 추가 시점 패치들이 포함되었다. 그림 9(b)는 상기 옵션에 0.5의 값을 할당 시 아틀라스를 도시한다. 모든 그룹의 첫 번째 아틀라스는 기본 시점만을 포함하고, 8개의 아틀라스에 총 16개의 기본 시점이 포함되었다. ‘Anchor\_atlas8’ 실험 조건에서는 상기 옵션에 0.5의 값을 할당 시 패킹 과정에서 오류가 발생하므로 상기 옵션 값을 0.4로 조정하여 최대 12개의 기본 시점을 아틀라스 내에 포함할 수 있으므로, 공정한 비교를 위해 ‘Group4\_atlas2’ 실험 조건에 대해서도 상기 옵션에 0.4의 값을 적용하였다.

본 논문은 [26]이 제안한 아틀라스 내 패치 점유율 출력 기능과 본 논문에서 제안하는 누락된 패치 비율 출력 기능을 통해 패치 누락을 방지하기 위한 아틀라스 개수 탐색을 위해 2, 4, 6, 8개의 아틀라스를 사용하여 실험을 진행하였다. 표 4는 상기 실험 조건에서의 아틀라스 내 패치 점유율 및 누락된 패치 비율을 나타낸다. ‘Anchor\_atlas2’ 실험 조건은 2개의 아틀라스 내에 패치들을 저장하므로 아틀라스

표 4. 아틀라스 내 패치 점유율 및 누락된 패치 비율  
 Table 4. Ratios of patch area in atlases and dropped patches

Condition	Ratio of patch area in atlases	Ratio of dropped patch area	Sum
Anchor_atlas2	200.00%	374.70%	574.70%
Anchor_atlas4	400.00%	126.13%	526.13%
Anchor_atlas6	563.76%	0.00%	563.76%
Anchor_atlas8	563.76%	0.00%	563.76%
Group2_atlas2	400.00%	138.24%	538.24%
Group4_atlas1	400.00%	140.47%	540.47%
Group3_atlas2	552.58%	9.01%	561.59%
Group6_atlas1	519.21%	0.00%	519.21%
Group4_atlas2	613.20%	0.00%	613.20%

내 패치 점유율은 200%를 초과할 수 없고, 아틀라스 내에 패킹 되지 못한 패치들은 누락된 패치 비율로 표현된다. ‘Anchor\_atlas2’ 실험 조건에서는 패치 비율의 총합이 574.70%였으므로 패치 누락이 발생하지 않으려면 최소 6개의 아틀라스가 필요함을 알 수 있다. ‘Anchor\_atlas4’ 실험 조건에서는 패치 비율의 총합이 526.13%로, ‘Anchor\_atlas2’ 대비 48.57%의 패치 비율 감소가 관찰되었다. 상기 감소에 대한 원인은 한 클러스터 내에 포함된 기본 시점의 개수가 증가함에 따라 프루닝 이후 남은 픽셀 수가 감소하였기 때문이다. ‘Anchor\_atlas2’ 실험 조건에서 두 개의 클러스터에 각각 두 개의 기본 시점이 포함되었던 반면, ‘Anchor\_atlas4’ 실험 조건에서는 세 개의 클러스터 중 두 개의 클러스터엔 세 개의 기본 시점이, 나머지 클러스터엔 두 개의 기본 시점이 포함되었다. ‘Anchor\_atlas2’ 실험 조건에서 처음으로 프루닝이 이뤄지는 추가 시점에서의 잔여 픽셀 수는 70% 수준으로 높은 반면, ‘Anchor\_atlas4’ 실험 조건에서는 30~40% 수준을 유지하여 패킹 되어야 할 픽셀 수 차이로 인해 상기 패치 비율이 차이가 났음을 알 수 있다. 반면 아틀라스 수를 4개에서 6개로 늘렸을 경우 37.63%의 패치 비율 증가가 관찰되었다. ‘Anchor\_atlas4’ 실험 조건에서 처음으로 프루닝이 이뤄지는 추가 시점에서의 잔여 픽셀 수는 30~40%인데 반해 ‘Anchor\_atlas6’ 실험 조건에서는 상기 픽셀 수가 10~20% 수준으로 유지되어 상기 원인은 본 실험 조건에서 큰 영향이 없었다. 따라서, 상기 현상의 원인 중 하나는 아틀라스 개수가 증가함에 따라

패킹 시 여유 공간이 증가하였고, 따라서 패치를 잘게 분할하지 않고도 삽입이 가능하였기 때문이다. 3.2 절에서 언급된 바와 같이 MIV 부호기는 패킹 시 패치가 이미 다른 패치가 점유중인 공간 내의 빈 공간에 삽입될 수 있는지 검사하고, 삽입이 불가하면 아틀라스 내에 다른 패치가 점유하지 않은 빈 공간에 패킹을 시도한다. 이 경우에도 삽입이 불가하면 패치는 더 작은 패치들로 분할되어 다시 패킹이 이루어진다. 아틀라스 개수가 적을 경우 아틀라스 내에 기존 패치들이 점유하지 않은 공간이 작고, 따라서 패치 내 빈 공간에 삽입을 시도하는 비율이 높다. 반면, 아틀라스 개수가 많을 경우 기존 패치들이 점유하지 않은 공간이 상대적으로 크므로 기존 패치 내에 삽입이 실패할 경우 아틀라스 내 빈 공간에 삽입되어 아틀라스 개수가 적을 때 대비 아틀라스 내 패치 점유율이 증가하게 된다. 패치 점유율이 이미 500% 후반에서 수렴하였기 때문에, 아틀라스 수를 6개에서 8개로 늘렸을 때 패치 비율의 변동이 관찰되지 않았다. 따라서, CBAbasketball 시퀀스에서는 패치 누락을 막기 위해 아틀라스 6개를 사용할 수 있음을 확인하였다. 그룹 기반 부호화 기법을 적용하였을 때도 패치 비율은 500% 수준에 수렴하였고, 4개의 그룹을 사용하고 각 그룹별로 아틀라스 2개가 할당되는 ‘Group4\_atlas2’ 실험 조건에서는 613.20%의 패치 비율이 관찰되었다. 아틀라스 수가 같고 그룹을 미적용한 ‘Anchor\_atlas8’ 실험 조건은 패치 패킹 시 첫 번째부터 여덟 번째 아틀라스 순서대로 패치 내 패치 패킹을 시도하여 첫 번째에 가까운 아틀라스일수록 작은 패치들의 수가 많을 확률이 높다. 반면, 그룹을 적용한 ‘Group4\_atlas2’ 실험 조건에서는 각 그룹별로 프루닝 및 패킹이 진행되어 2개의 아틀라스 내에서 패치 패킹 수행이 시도되는데, 8개의 아틀라스를 순회하며 패치 내 빈 공간을 찾아 삽입하는 ‘Anchor\_atlas8’ 실험 조건 대비 패치 내 패치 패킹보다는 아틀라스 내 빈 공간에 삽입되는 횟수가 더 많아 프루닝 후 잔여 픽셀 수와 별개로 아틀라스 내 패치 비율이 상대적으로 높았음을 알 수 있다. 결과적으로 아틀라스 및 그룹 개수에 따른 차이는 있지만 본 실험에 사용된 CBAbasketball 시퀀스의 경우 아틀라스 6개를 사용 시 패치 누락이 일어나지 않음을 알 수 있었고, 이는 본 논문이 제안한 패치 비율 출력 기능을 통해 유추가 가능하였다.

표 5. 그룹 기반 MIV의 PSNR 및 IV-PSNR BD-rate (음수 값이 BD-rate 절감을 나타냄)

Table 5. PSNR and IV-PSNR BD-rates of group-based MIV compared to the anchor (Negative value indicates bitrate saving)

	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
Group2_atlas2	-30.09%	-21.14%	-38.27%	-26.04%
Group4_atlas1	61.57%	30.44%	33.80%	19.36%
Group3_atlas2	3.24%	1.67%	-2.02%	0.07%
Group6_atlas1	14.18%	10.35%	18.28%	13.84%
Group4_atlas2	-2.60%	0.71%	-0.16%	2.86%
Average	9.26%	4.41%	2.33%	2.02%

표 5는 그룹을 적용하지 않은 MIV 대비 그룹을 적용 시 PSNR 및 IV-PSNR에 대한 Bjøntegaard-delta rate (BD-rate) 값을 나타낸다. BD-rate는 4개의 비트율 지점을 기반으로 산출되고, MIV CTC는 5개의 비트율 지점에 대해 실험을 진행할 것을 권장하여 상위 4개의 비트율 지점에 대해 high BD-rate, 하위 4개의 비트율 지점에 대해 low-BD-rate 라는 명칭을 사용한다. 표 4에 도시된 바와 같이, CBA-basketball 시퀀스는 4개의 아틀라스를 사용하여 MIV 부호화 시 일부 패치들이 누락되어 해당 패치가 소속된 시점을 합성 시 품질 저하가 발생할 수 있다. 또한, MIV 부호화가 기본 시점들을 먼저 아틀라스에 패킹하고 그 이후 추가 시점 패치들을 패킹함을 고려했을 때, 입력 영상의 크기 및 개수 대비 적은 수의 아틀라스를 사용 시 다수의 기본 시점들이 패킹되는데 반해 추가 시점 패치들의 누락이 발생할 수 있다. 따라서, 제한된 수의 아틀라스를 사용하여 MIV 부호화 시 그룹 개수를 늘리는 것 보다 그룹 별 아틀라스 수를 늘리는 것이 몰입형 영상에 대한 전역적 표현에 유리할 수 있다. ‘Group4\_atlas1’ 실험 조건은 4개의 그룹 별로 하나의 아틀라스를 할당하여 추가 시점 패치들의 손실이 발생하였고, 따라서 그룹 미적용 실험 조건인 ‘Anchor\_atlas4’ 대비 BD-rate 손실이 발생하였다. 반면 ‘Group2\_atlas2’ 실험 조건은 ‘Group4\_atlas1’ 실험 조건 대비 적은 수의 기본 시점들을 포함했으나 그룹 당 2개의 아틀라스를 사용하여 추가 시점 패치들을 잘 보존하였고, 평균 25.61%, 32.15%의 PSNR 및 IV-PSNR BD-rate 절감을 기록하였다. ‘Group3\_atlas2’ 및 ‘Group6\_atlas1’ 실험 조건을 비교했을

때에도 그룹 수는 적지만 그룹 별 아틀라스 수가 더 많은 ‘Group3\_atlas2’ 실험 조건이 좀 더 우수한 BD-rate를 보여 주었다. 아틀라스 8개를 출력하는 ‘Group4\_atlas2’ 실험 조건은 그룹을 적용하지 않은 ‘Anchor\_atlas8’ 조건 대비 높은 대역폭에서 BD-rate 절감을 보여주었다.

그림 10은 본 실험의 율-왜곡 곡선 (rate-distortion curve, RD-curve) 을 도시한다. 그림 10(a)에 도시된 바와 같이, CBAbasketball 시퀀스는 아틀라스 6개를 사용하는 실험 조건에서 수렴하였고, 따라서 아틀라스 8개를 사용하여도 율-왜곡 곡선은 아틀라스 6개를 사용했을 때와 동일하였다. 그림 10(b)는 그룹 기반 부호화 실험 조건들의 율-왜곡 곡선을 도시한다. 그룹 별 아틀라스를 충분히 확보하지 못한 ‘Group4\_atlas1’ 실험 조건이 제일 낮은 효율을 보였고, 그룹 수가 적은 대신 그룹 별 아틀라스를 2개 확보한 ‘Group2\_atlas2’ 실험 조건이 상기 실험 조건보다 좋은 효율을 보였다. 표 4에 도시된 바와 같이 CBAbasketball 시퀀스는 아틀라스 내 패치 점유율이 500% 후반에 수렴하였으므로 아틀라스를 6개 사용할 때는 패치 누락이 일어나지 않아 ‘Group6\_atlas1’ 실험 조건은 ‘Group2\_atlas2’ 실험 조건 대비 좋은 효율을 기록하였다. 동일하게 아틀라스를 6개 사용하지만 그룹을 적게 분할하고 그룹 별 아틀라스를 더 많이 확보한 ‘Group3\_atlas2’ 실험 조건은 ‘Group6\_atlas1’ 실험 조건 대비 좋은 효율을 보여주었고, 그룹을 4개로 분할하고 그룹 별 2개의 아틀라스를 할당한 ‘Group4\_atlas2’ 실험 조건은 저대역폭에서는 ‘Group3\_atlas2’ 실험 조건 대비 낮은 효율을 보였으나 고대역폭에서는 높은 효율을 보였다. 두 조건 모두 패치가 누락되지 않을 정도의 아틀라스 수를 확보했으나, 그룹을 더 많이 분할하여 그룹 별 시점 수가 상대적으로 적었던 ‘Group4\_atlas2’ 실험 조건이 ‘Group3\_atlas2’ 실험 조건 대비 많은 양의 추가 시점 패치들을 생성하였고, 이로 인해 고대역폭에서는 ‘Group4\_atlas2’ 실험 조건이 시점 합성 시 유리했으나 저대역폭에서는 비트율이 적게 드는 ‘Group3\_atlas2’ 실험 조건이 좋은 결과를 보여주었다. 그림 10(c), 10(d)는 아틀라스 4개 사용 시 그룹 기반 MIV의 율-왜곡 곡선을 도시하고, 상기 그림을 통해 패치 누락이 발생하는 환경에서는 그룹 별 아틀라스 개수를 충분히 확보하는 것이 유리함을 알 수 있다. 그림 10(e), 10(f) 역시 그룹 별 아틀라스 개수를 늘린 실험 조건

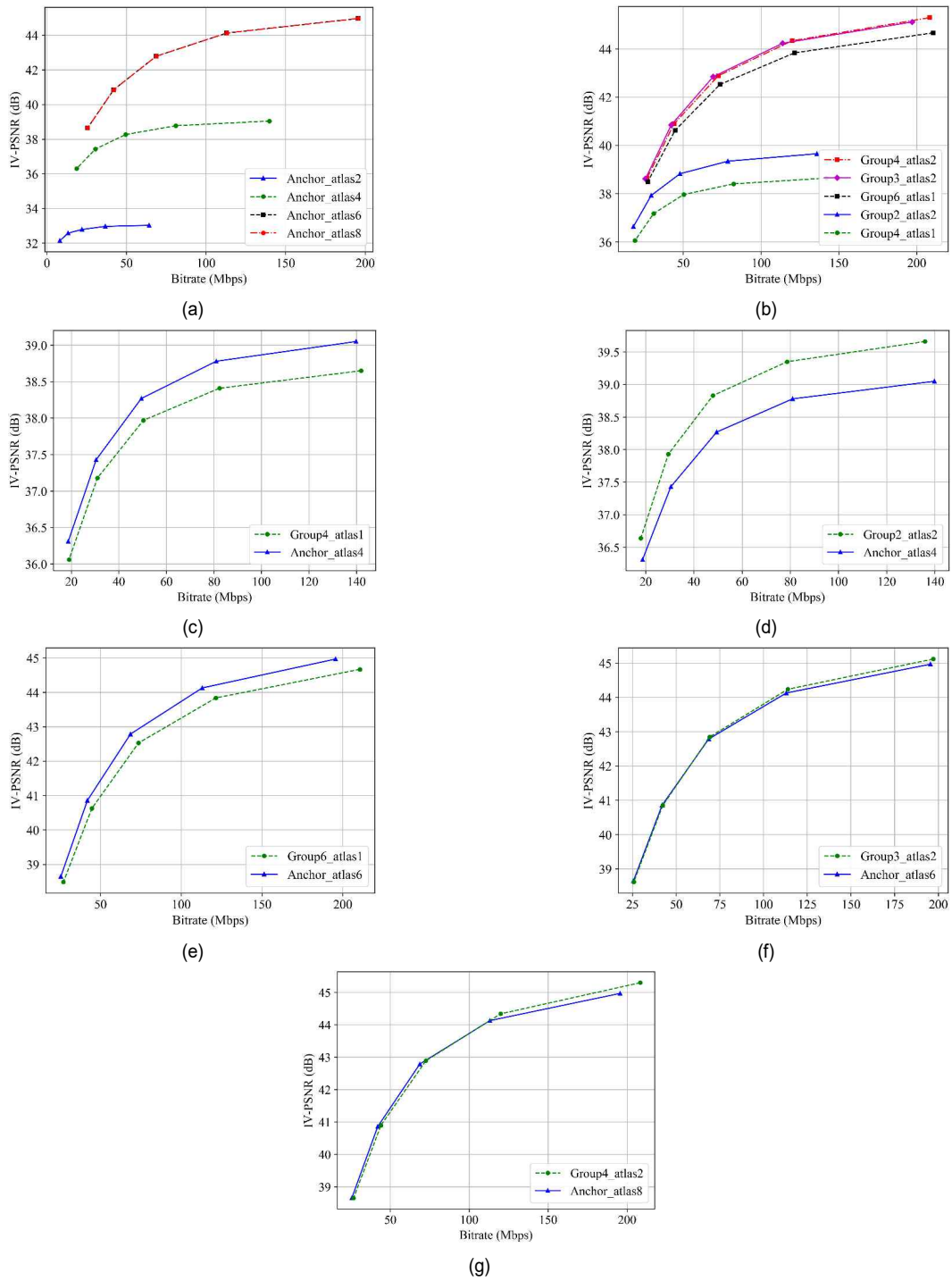


그림 10. 울-왜곡 곡선. (a) 그룹 미적용 부호화 기법, (b) 그룹 적용 부호화 기법, (c) 'Group4\_atlas1' 실험 조건, (d) 'Group2\_atlas2' 실험 조건, (e) 'Group6\_atlas1' 실험 조건, (f) 'Group3\_atlas2' 실험 조건, (g) 'Group4\_atlas2' 실험 조건.

Fig. 10. Rate-distortion curves. (a) non group-based methods, (b) group-based methods, (c) 'Group4\_atlas1', (d) 'Group2\_atlas2', (e) 'Group6\_atlas1', (f) 'Group3\_atlas2', (g) 'Group4\_atlas2'.

이 기존 그룹 미적용 기법 대비 좋은 효율을 보임을 증명하였다. 그림 10(g)를 통해 패치 누락이 발생하지 않을 정도로 많은 아틀라스를 사용 가능할 경우 그룹 기반 MIV가 그룹 미적용 대비 고대역폭에서 좋은 효율을 보임을 알 수 있다.

본 논문은 표 5와 그림 10에 의해 증명된 그룹 기반 MIV의 화질 우수성을 주관적 측면에서 검증하기 위해 합성된

가상 시점 별 비교를 진행하였다. 그림 11은 그룹 적용/미적용 시 QP4 지점에서 합성된 중간 시점에서 주관적 화질 차이를 보여주는 지점을 확대한 그림을 도시한다. 그림에서 주관적 화질 차이가 두드러지는 부분은 빨강 점선으로 표시되었다. 그림 11(a)는 ‘Anchor\_atlas4’ 실험 조건에서의 시점 v09 위치에서 합성된 중간 시점을 나타내고, 물체

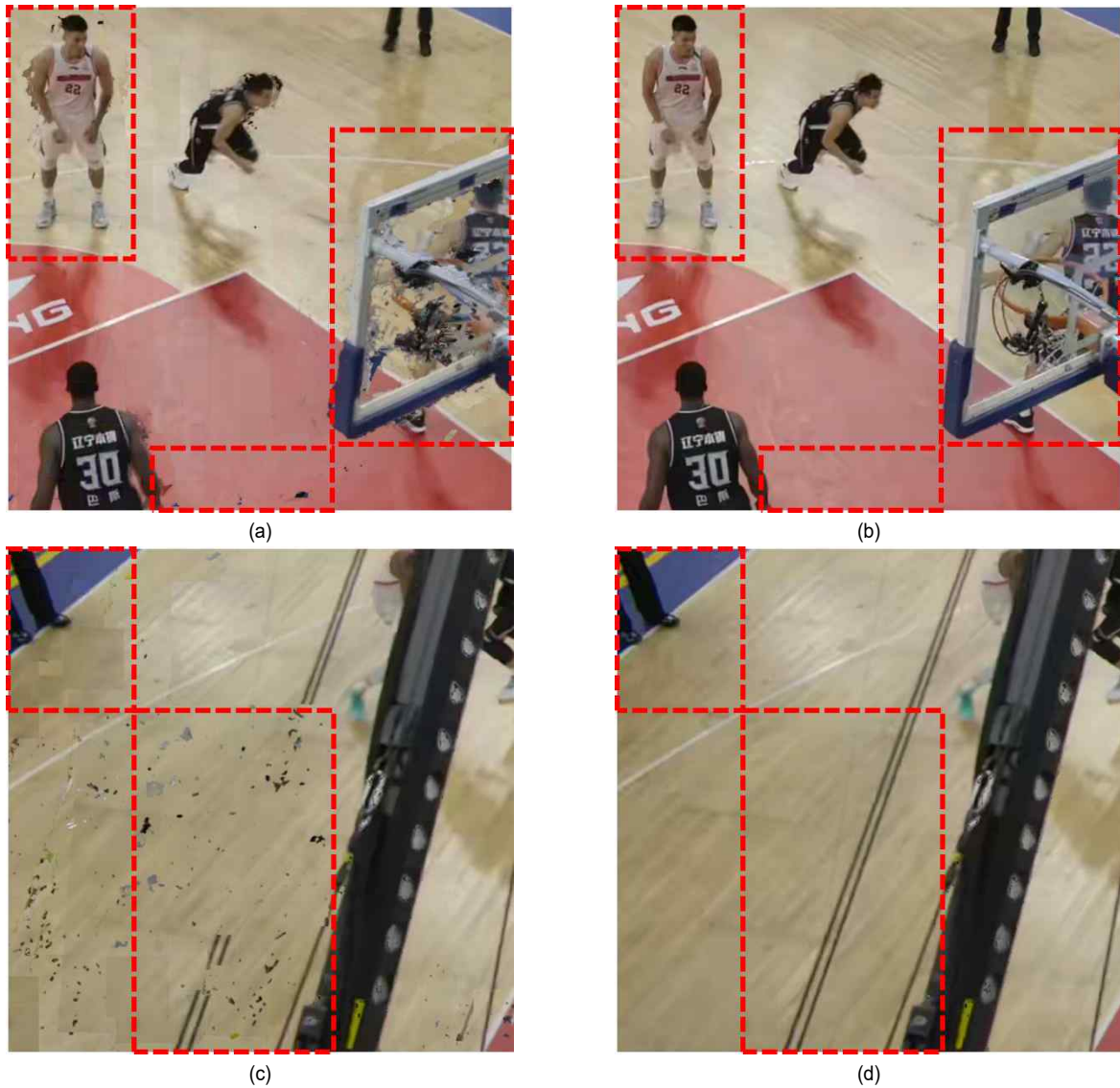


그림 11. 그룹 적용/미적용 시 QP4 지점에서 합성된 중간 시점 확대. (a) Anchor\_atlas4, view v09, 30.42Mbps, (b) Group2\_atlas2, view v09, 29.38Mbps, (c) Anchor\_atlas6, view v19, 41.95Mbps, (d) Group3\_atlas2, view v19, 42.28Mbps.

Fig 11. Enlarged synthesized intermediate views of group-based and non-group-based methods at QP4, (a) Anchor\_atlas4, view v09, 30.42Mbps, (b) Group2\_atlas2, view v09, 29.38Mbps, (c) Anchor\_atlas6, view v19, 41.95Mbps, (d) Group3\_atlas2, view v19, 42.28Mbps.

경계 및 내부에서 합성 오류에 의한 아티팩트가 다수 발생하였다. 반면 ‘Group2\_atlas2’ 실험 조건 사용 시 그림 11(b)에 도시된 바와 같이 상기 아티팩트가 상당 부분 제거되었다. 그림 11(c)는 ‘Anchor\_atlas6’ 실험 조건에서의 시점 v19 위치에서 합성된 중간 시점을 나타낸다. 바닥에서 다수의 블로킹 아티팩트 (blocking artifact) 및 합성 오류가 발생하였으나, ‘Group3\_atlas2’ 실험 조건 사용 시 그림 11(d)에 도시된 바와 같이 화질이 상당 부분 개선되었음을 확인하였다.

표 5와 그림 10에 도시된 실험 결과에 따르면 그룹 기반 MIV는 그룹 미적용 대비 BD-rate 손실이 다소 발생한다. 그러나, 그림 11에 도시된 바와 같이 주관적 화질 면에서 그룹 기반 MIV의 화질 개선을 확인하였고, 아틀라스 수가 적어 패치 누락이 발생할 경우 그룹 별 아틀라스 수를 충분히 확보할 경우 그룹 미적용 대비 높은 효율을 보일 수 있으며, 사용자 시점 기반 스트리밍 시 그룹을 사용할 경우 사용자 관심 영역에 해당하는 그룹만 전송할 수 있으므로 그룹 미적용 대비 대역폭 절감이 가능하다. 또한 패치 누락에 의한 가상 시점 합성 시 품질 하락을 방지하기 위해 본 논문이 제안하는 패치 점유율 및 누락된 패치 비율 출력 기능을 통해 시퀀스별 아틀라스 수를 결정할 수 있다.

## V. 결론

본 논문은 대형 공간을 취득한 몰입형 영상을 효율적으로 압축하기 위한 그룹 기반 MIV 및 패치 점유율, 누락된 패치 비율 출력 기능을 최신 버전의 TMIV에서 구현하였다. 제안하는 구현을 통해 몰입형 영상은 그룹 단위로 분할되어 프루닝 및 패키징이 진행되고, 그룹 별 메타데이터는 통합되어 단일 비트스트림 형태로 저장된다. MIV CTC가 권장하는 2개의 아틀라스를 사용 시 아틀라스 내 공간 부족으로 인해 패치 누락이 발생할 경우엔 그룹을 많이 분할하기보다는 그룹 별 아틀라스 수를 충분히 확보하는 것이 효율적임을 확인하였고, 로그에 출력된 패치 정보를 통해 입력된 몰입형 영상에 적합한 아틀라스 수를 결정할 수 있음을 확인하였다. 향후 연구로는 특정 그룹의 메타데이터만 추출하여 부분 접근 (partial access) 이 가능하게 하고 이를 통해 6DoF 사용자 시점 기반 스트리밍을 연구 및 구현할 계획이다.

## 참고 문헌 (References)

- [1] E. S. Jang, M. Preda, K. Mammou, A. M. Tourapis, J. Kim, D. B. Graziosi, S. Rhyu, M. Budagavi. 2019. Video-based point-cloud-compression standard in mpeg: From evidence collection to committee draft [standards in a nutshell]. *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 118-123. doi: <https://doi.org/10.1109/MSP.2019.2900721>
- [2] S. Schwarz, M. Pesonen. 2019. Real-time decoding and AR playback of the emerging MPEG video-based point cloud compression standard. *Nokia Technologies; IBC: Helsinki, Finland*.
- [3] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, L. Yu. 2021. MPEG immersive video coding standard. *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1521-1536. doi: <https://doi.org/10.1109/JPROC.2021.3062590>
- [4] J. -B. Jeong, S. Lee, E. -S. Ryu. 2022. Rethinking Fatigue-Aware 6DoF Video Streaming: Focusing on MPEG Immersive Video. *International Conference on Information Networking 2022 (ICOIN2022)*, pp. 304-309. doi: <https://doi.org/10.1109/ICOIN53446.2022.9687247>
- [5] Sangho Lee, Hongchang Shin, Gwangsoon Lee, Jeongil Seo. 2022. GPU Implementation of TMIV Decoder for Real-time Playback. *The Korean Institute of Broadcast and Media Engineers (KIBME) Summer Conference*, pp. 93-96.
- [6] G. J. Sullivan, J. -R. Ohm, W. -J. Han, T. Wiegand. 2012. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649-1668. doi: <https://doi.org/10.1109/TCSVT.2012.2221191>
- [7] B. Bross, Y. -K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, J. -R. Ohm. 2021. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 10, pp. 3736-3764. doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [8] Y. Bai, X. Sheng, S. Li, C. Wang, L. Yu. 2021. [MIV]Undistorted CBA Basketball Test Sequence for MPEG-I Visual, Standard ISO/IEC JTC1/SC29/WG4, m58500.
- [9] ISO/IEC JTC 1/SC 29/WG 7. 2021. Text of ISO/IEC DIS 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition. Standard ISO/IEC JTC 1/SC 29 WG 7, n00188.
- [10] ISO/IEC JTC 1/SC 29/WG 4. 2022. Preliminary WD1 of ISO/IEC 23090-12 MPEG Immersive video Ed. 2. Standard ISO/IEC JTC 1/SC 29/WG 4, n00176.
- [11] D. Park, S. -G. Lim, K. -J. Oh, G. Lee, J. -G. Kim. 2022. Nonlinear Depth Quantization Using Piecewise Linear Scaling for Immersive Video Coding. *IEEE Access*, Vol. 10, pp. 4483-4494. doi: <https://doi.org/10.1109/ACCESS.2022.3140537>
- [12] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. Sub-bitstream packing based lightweight tiled streaming for 6 degree of freedom immersive video. *Electronics Letters*, Vol. 57, No. 25, pp. 973-976. doi: <https://doi.org/10.1049/ell2.12329>

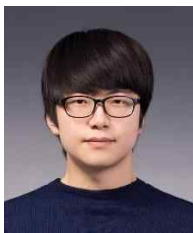


- [13] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. DWS-BEAM: Decoder-Wise Subpicture Bitstream Extracting and Merging for MPEG Immersive Video. International Conference on Visual Communications and Image Processing 2021 (VCIP2021), pp. 1-5.  
doi: <https://doi.org/10.1109/VCIP53242.2021.9675419>
- [14] H. -C. Shin, J. -Y. Jeong, G. Lee, M. U. Kakli, J. Yun, J. Seo. 2022. Enhanced pruning algorithm for improving visual quality in MPEG immersive video. ETRI Journal, Vol. 44, No. 1, pp. 73-84.  
doi: <https://doi.org/10.4218/etrij.2021-0211>
- [15] S. Kwak, J. Yun, J. -Y. Jeong, Y. Kim, I. Ihm, W. -S. Cheong, J. Seo. 2022. View synthesis with sparse light field for 6DoF immersive video. ETRI Journal, Vol. 44, No. 1, pp. 24-37.  
doi: <https://doi.org/10.4218/etrij.2021-0205>
- [16] B. Salahieh, S. Bhatia, J. Boyce. 2019. Group-based TMIV. Standard ISO/IEC JTC1/SC29/WG11, MPEG/m49406.
- [17] S. Lee, J. -B. Jeong, E. -S. Ryu. 2022. Group-based Adaptive Rendering for 6DoF Immersive Video Streaming. Journal of Broadcast Engineering, Vol. 27, No. 2, pp. 216-227.  
doi: <https://doi.org/10.5909/JBE.2022.27.2.216>
- [18] S. Lee, J. -B. Jeong, E. -S. Ryu. 2022. Efficient Group-Based Packing Strategy for 6DoF Immersive Video Streaming. International Conference on Information Networking 2022 (ICOIN2022), pp. 310-314.  
doi: <https://doi.org/10.1109/ICOIN53446.2022.9687139>
- [19] B. Salahieh, B. Kroon, J. Jung, M. Domański. 2019. Test Model for Immersive Video. Standard ISO/IEC JTC 1/SC 29 WG 11, MPEG/n18470.
- [20] V. K. M. Vadakital. 2019. Description of Immersive Video Core Experiments 1. Standard ISO/IEC JTC1/SC29/WG11, MPEG/n18465.
- [21] B. Salahieh, J. Fleureau, R. Dore, J. Boyce. 2020. Interactions Between Grouping, Pruning, and Synthesizing. Standard ISO/IEC JTC 1/SC 29 WG 11, MPEG2020/m52824.
- [22] T. T. Le, J. -B. Jeong, S. S. Lee, J. Kim, E. -S. Ryu. 2021. An Efficient Viewport-Dependent 360 VR System Based on Adaptive Tiled Streaming. Computers, Materials & Continua, Vol. 66, No. 3, pp. 2627-2643.  
doi: <https://doi.org/10.32604/cmc.2021.013399>
- [23] J. -B. Jeong, S. Lee, I. Kim, E. -S. Ryu. 2021. Implementing Viewport Tile Extractor for Viewport-Adaptive 360-Degree Video Tiled Streaming. International Conference on Information Networking 2021 (ICOIN2021), pp. 8-12.  
doi: <https://doi.org/10.1109/ICOIN50884.2021.9333964>
- [24] B. Sonneveldt, B. Kroon. 2021. Philips CE-2.10 response “Server-side Inpainting 2”. Standard ISO/IEC JTC 1/SC 29 WG 4, MPEG/m55736.
- [25] J. Jung, B. Kroon. 2022. Common Test Conditions for MPEG Immersive Video. Standard ISO/IEC JTC 1/SC 29/WG 4, n00232.
- [26] K. -J. Oh, J. Y. Oh, T. Dong, X. Li, G. Lee, E. S. Jang. 2022. [EE-7 related] Atlas cropping-based TMIV encoding results. Standard ISO/IEC JTC 1/SC 29/WG 4, m60274.
- [27] A. Dziembowski, D. Mieloch, J. Stankowski, A. Grzelka. 2022. IV-PSNR – The objective quality metric for immersive video applications. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 32, No. 11, pp. 7575-7591.  
doi: <https://doi.org/10.1109/TCSVT.2022.3179575>

---

저 자 소 개

---



**정 종 범**

- 2018년 8월 : 가천대학교 컴퓨터공학과 학사
- 2018년 9월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육학과 석박통합과정
- 2020년 1월 ~ 2020년 3월 : University of California, Santa Barbara 방문연구원
- 2021년 8월 ~ 2022년 1월 : Purdue University 방문연구원
- 2022년 9월 ~ 현재 : 성균관대학교 글로벌융합학부 강사
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 관심분야: 멀티미디어 통신 및 시스템, 비디오 압축 표준, MPEG immersive video, video-based dynamic mesh coding

---

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, MPEG immersive video, Deep learning



최 재 열

- 2018년 3월 ~ 현재 : 성균관대학교 컴퓨터교육과 재학
- ORCID : <https://orcid.org/0009-0009-2923-1252>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준, MPEG immersive video, Deep learning



이 광 순

- 1993년 2월 : 경북대학교 전자공학과 학사
- 1995년 8월 : 경북대학교 전자공학과 석사
- 2004년 8월 : 경북대학교 전자공학과 박사
- 2001년 2월 ~ 현재 : 한국전자통신연구원 미디어연구본부 책임연구원
- 주관심분야 : MPEG Immersive video 복호화, Implicit Neural Visual Representation 기반 입체영상 렌더링 기술



곽 상 운

- 2012년 2월 : 연세대학교 전기전자공학과 학사
- 2014년 2월 : 한국과학기술원 전기 및 전자공학과 석사
- 2014년 3월 ~ 현재 : 한국전자통신연구원 미디어연구본부 선임연구원
- ORCID : <https://orcid.org/0000-0002-9295-8966>
- 주관심분야 : 이머시브 미디어 기술, 기계를 위한 영상 부호화 및 국제 표준, 딥러닝 기반 장면 표현 기술



정 원 식

- 1992년 2월 : 경북대학교 전자공학과 (공학사)
- 1994년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 2000년 2월 : 경북대학교 대학원 전자공학과 (공학박사)
- 2000년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0001-5430-2969>
- 주관심분야 : 3DTV 방송 시스템, 라이트필드 이미징, 영상부호화, 딥러닝 기반 신호처리, 멀티미디어 표준화

---

저 자 소 개

---



**이 봉 호**

- 1997년 2월 : 한국항공대학교 전자공학과 학사
- 1999년 2월 : 한국항공대학교 전자공학과 공학석사
- 1999년 7월 ~ 현재 : 한국전자통신연구원 미디어연구본부 책임연구원
- ORCID : <http://orcid.org/0000-0002-4502-1361>
- 주관심분야 : 방송 시스템, 모바일 방송, 이머시브 미디어 기술



**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