MPEG Immersive Video 를 위한 그룹 기반 적응적 스트리밍

정종범 [1], 이순빈 [1], 최재열 [1], 이광순 [2], 곽상운 [2], 정원식 [2], 이봉호 [2], 류은석 [1]
1) 성균관대학교 컴퓨터교육학과, 2) 한국전자통신연구원
{uof4949, soonbinlee, jaychoi}@skku.edu, {gslee, s.kwak, wscheong, leebh}@etri.re.kr,
esryu@skku.edu

Towards Group-based Adaptive Streaming for MPEG Immersive Video

Jong-Beom Jeong, Soonbin Lee, Jaeyeol Choi, Gwangsoon Lee, Sangwoon Kwak, Won-Sik Cheong, Bongho Lee, Eun-Seok Ryu
1) Department of Computer Science Education, Sungkyunkwan University
2) Electronics and Telecommunications Research Institute

요 약

다수의 색상 및 거리 정보로 구성된 몰입형 영상 부호화를 위한 MPEG immersive video (MIV) 표준은 각 시점의 영상 간 중복성 제거 및 잔여 영상 병합을 통한 압축률 향상을 목표로 한다. 시점에 따른 카메라 그룹핑을 통해 압축률 향상이 가능하나, 그룹 기반 MIV 부호화 기술은 최근 활발히 논의되고 있지 않다. 따라서 본 논문은 최신 버전의 MIV 참조 소프트웨어에 그룹 기반 부호화 기술을 이식하고 적응적 스트리밍을 위한 그룹 기반 부호화 기술의 효율을 검증하였다.

1. 서론

몰입형 메타버스 환경이 주목받으면서 더욱 생동감있는 가상 및 증강 현실을 표현할 수 있는 부호화, 전송, 렌더링 기술들이 주목받고 있다. 3 차원 공간을 나타내는 3-D 포인트 클라우드 (point cloud) 를 2 차원 공간에 매핑 (mapping) 후 기존 2-D 영상 압축 표준 (e.g., high-efficiency video coding (HEVC)) 을 사용하여 높은 압축률을 확보하는 video-based point cloud coding (V-PCC) 기술이 moving picture experts group (MPEG) 에 의해 표준화가 진행되고 있으며, Nokia 에 의해 실시간 구현이 가능함이 증명되었다[1, 2]. 한편, 다수의 2-D 카메라 배열을 통해 취득된, 텍스쳐 (색상) 와 지오메트리 (거리) 순서쌍으로 이루어진 몰입형 영상을 압축하는 MPEG immersive video (MIV) 표준 역시 MPEG 에 의해 표준화가 진행 중에 있고, 현재 2nd edition 을 위한 논의가 진행되고 있으며, V-PCC 와 마찬가지로 실시간화가 가능함이 증명되었다[3, 4]. MIV 는 다수의 몰입형 영상 및 카메라 매개변수를 입력 받아 영상 간 중복성 제거 및 잔여 영상 병합을 통해 아틀라스 (atlas) 라 불리는, 입력 몰입형 영상 대비 비트율, 복호기 개수를 크게 절약한 영상을 출력하고, 이를 HEVC, versatile video coding (VVC) 등의 2-D 영상 부호화 표준을 사용하여 높은 부호화 효율을 보인다.

MIV 부호기는 복호기 단 뷰 합성을 고려하지 않고 압축률을 높이기 위해 시점 간 중복성을 제거하는 프루닝 (pruning) 을 수행하므로 비트율 대비 품질이 하락하는 사례가 일부 발견되었다. 이 문제를 해결하기 위해 MIV 그룹 내에서 그룹 기반 부호화 기법이 제안되었다[5]. 입력된 몰입형 영상들은 비슷한 시점들끼리 그룹으로 묶이고, 이후 MIV 부호기는 프루닝 및 잔여 영상을 병합하는 패킹 (packing) 작업을 그룹 단위로 수행한다. 이를 통해 각 그룹에서 중요한 잔여 영상이 보존되고, 합성된 뷰 품질이 향상되어 율-왜곡 (rate-distortion) 측면에서의 최적화가 가능하다[6, 7]. 하지만, 그룹 기반 MIV 의 효율에도 불구하고 그룹 기반 부호화 기술이 MIV 참조 소프트웨어인 test model for immersive video (TMIV) 에 적용되면서 소스 코드 관리가 어려워져 MIV 표준화 그룹은 그룹 기반 부호화 기술을 TMIV 에서 제외하였고, 최신 버전의 TMIV 에서는 그룹 기반 부호화 및 그에 대한 성능 평가가 이루어지고 있지 않다[8]. 그러나, 그룹 기반 MIV 부호화 기술은 그림 1 에 도시된 대규모 실사 컨텐츠 부호화 시 유효하게 사용될 수 있으며, 이에 대한 성능 평가가 필요하다.

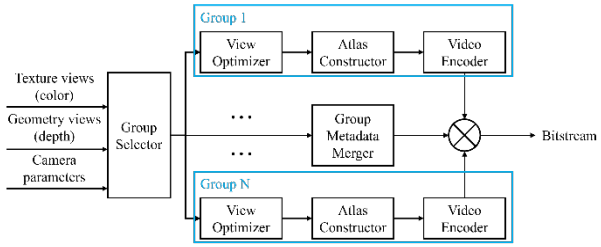


그림 1. 그룹 기반 MIV 부호화 블록 다이어그램.

본 논문은 MIV 에서의 그룹 기반 부호화 기술을 최신 버전의 TMIV 에서 구현하고, 실사 컨텐츠에 대한 그룹 기반 MIV 부호화 기술의 효율을 검증한다. 그림 1 은 그룹 기반 MIV 부호기의 동작을 도시한다. 그룹 셀렉터는 카메라 매개변수를 기반으로 텍스쳐와 지오메트리로 이루어진 몰입형 영상을 다수의 그룹으로 분할한다. 이후 view optimizer 는 영상의 위치 및 시야각을 고려하여 몰입형 영상을 기본 시점 (basic view) 및 추가 시점 (additional view) 로 분류하고, atlas constructor 는 시점 간 중복성을 추가 시점에서 제거하고 잔여 영상을 아틀라스에 저장하며 상기 과정들은 그룹 단위로 수행된다. HEVC, VVC 등의 영상 부호기를 통해 아틀라스들이 압축되어 영상 비트스트림들이 생성되고, 상기 비트스트림들은 아틀라스를 통한 뷰 합성 시 필요한 메타데이터들을 담은 MIV 비트스트림과 멀티플렉싱 (multiplexing) 되어 단일 비트스트림이 생성된다.

본 논문의 구성은 다음과 같다. 2 절에서는 그룹 기반 MIV 의 배경 및 관련 연구를 소개한다. 3 절에서는 그룹 기반 MIV 의 구현 내용 및 실험 결과를 설명한다. 마지막으로 4 절에서는 본 논문의 결론을 맺는다.

2. 배경 및 관련 연구

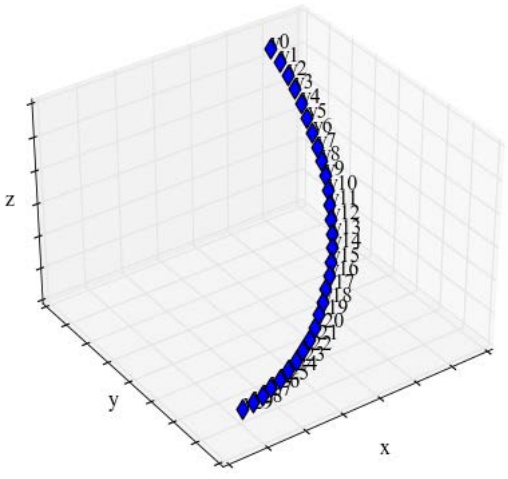
2019 년 5 월에 TMIV v1.0 릴리즈 이후, MIV 표준화 그룹 내에서는 MIV 의 성능 향상을 위한 core experiment (CE) 및 exploration experiment (EE) 를 진행하였다. 2019 년에 진행된 CE1 의 요구사항인 적응적 입력 영상/가상 시점 영상 선택 및 처리를 만족시키기 위해[9], Intel 은 그룹 기반 MIV 를 제안하였다[10]. MIV 부호기는 제한된 아틀라스 내에 최대한 많은 잔여 영상을 직사각형 형태의 패치 (patch) 로 삽입하므로 MIV 부호기가 가상 시점 영상 합성 시 배경을 나타내는 패치가 전경 물체 위에 렌더링되어 합성된 시점 품질이 저하되는 문제가 발생하였다. 그룹 기반 MIV 는 그룹 단위로 기본/추가 시점 분류, 프루닝, 패킹을 진행하여 일관된 지역적 특성을 잘 보존하고 합성된 시점 품질을 크게 향상하였다. 그룹 기반 MIV 부호화 기법은 TMIV v3.0 에 채택되었으나, TMIV v9.0 이후로는 소스 코드 관리의 어려움으로 인해 TMIV 에서 제외되었다.

그룹 기반 MIV 는 선택적 스트리밍 구현이 용이하다는 또 다른 장점을 가진다. 대형 실사 컨텐츠를 MIV 부호화 후 전송할 때, 그룹을 사용하지 않았을 시에는 사용자 시점을 표현하는 아틀라스 및 MIV 비트스트림을 선택적으로 전송하기 어렵지만, 그룹 사용 시 특정 그룹의 아틀라스 및 MIV 비트스트림을 추출하여 전송하는 기능의 구현이 상대적으로 용이하다. 서브픽쳐 기반 선택적 MIV 스트리밍 기술과 함께 사용되면 비트율 절약 및 고품질 스트리밍이 가능하므로, 최신 버전의 TMIV 에서 그룹 기반 MIV 부호화 기술 효율 검증이 논의될 수 있다[11, 12].

3. 그룹 기반 MIV 구현 및 실험 결과

본 절은 최신 버전의 TMIV 에서 구현한 그룹 기반 MIV 부호화 기법 및 성능 평가에 대해 설명한다. 그룹 기반 MIV 부호화 기법은 TMIV v13.0 버전 상에서 구현되었다. TMIV v13.0 은 v11.0 대비 부호화/복호화 함수 호출 방식, 메타데이터를 표현하는 visual volumetric video coding (V3C) 구조체 구현 방식에서 상당한 변화가 있었고, 최근 릴리즈된 버전과 v13.0 간 차이는 미미하므로 안정성 역시 고려하여 v13.0 에서의 구현을 진행하였다. 성능 평가는 30 개의 2-D 카메라를 통해 취득된 CBAbasketball 테스트 시퀀스에 대해 진행되었다[13]. 그림 2 에 도시된 CBAbasketball 은 2048×1088 해상도의 카메라를 통해 농구 경기장 전역을 표현하였고, MIV 1st edition 에서 다뤄졌던 테스트 시퀀스에 비해 넓은 움직임 시차 (motion parallax) 를 제공해야 하므로 본 실험에 사용되었다.

(a)



(b)

그림 2. 대규모 실사 컨텐츠 취득 예시 (sequence CBAbasketball). (a) 텍스쳐 영상, (b) 카메라 배열.

표 1. 그룹 기반 MIV 부호화 기법 BD-rate (음수가 비트율 절감을 나타냄)

Anchor	Proposed	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
Anchor (atlas4)	Group2 (atlas4)	-30.09%	-21.14%	-38.27%	-26.04%
	Group4 (atlas4)	61.57%	30.45%	33.80%	19.36%
Anchor (atlas6)	Group6 (atlas6)	14.18%	10.35%	8.68%	8.11%
Anchor (atlas8)	Group4 (atlas8)	-2.60%	0.71%	-7.52%	-2.10%

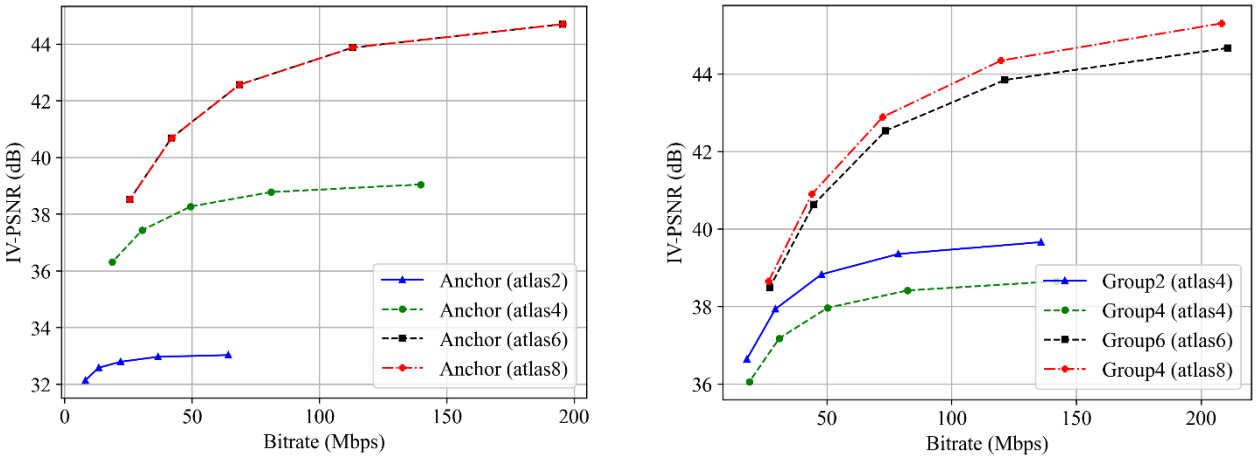


그림 3. 율-왜곡 곡선. (a) 그룹을 사용하지 않은 MIV 부호화 기법, (b) 그룹 기반 MIV 부호화 기법.

HEVC 에서의 부호화 효율을 검증하기 위해 HEVC test model (HM) 16.16 버전을 사용하여 아틀라스 부호화 및 복호화를 진행하였고, 영상 품질 평가 기법으로는 MIV 의 공통 실험 조건 (common test conditions; CTC) 에서 권고하는 immersive video peak signal-to-noise ratio (IV-PSNR) 및 Y-PSNR 이 채택되었다[14, 15]. Bjontegaard-delta rate (BD-rate) 기반 비트율 효율 검증을 위해 텍스쳐에 대해 22, 27, 32, 37, 42 의 양자화 매개변수 (quantization parameter; QP) 를 사용하였고, 지오메트리에 대해서는 텍스쳐 QP 를 MIV CTC 에서 제공하는 선형 방정식을 통해 매핑한 값인 3, 7, 11, 15, 19 의 QP 값을 사용하였다. 기존 MIV CTC 는 약 4K 의 해상도를 가지는 2 개의 텍스쳐 및 지오메트리 아틀라스를 생성할 것을 권고하나, 대형 실사 컨텐츠인 CBAbasketball 에서는 2 개의 아틀라스만을 사용했을 때 다수의 홀 (hole) 이 발생함을 고려하여 4, 6, 8 개의 아틀라스를 사용하여 실험을 진행하였다. 그룹 기반 MIV 부호화 기법은 그룹을 2 개, 4 개, 6 개 사용 시 그룹 당 아틀라스를 1 개 또는 2 개를 할당하여 진행하였다. 표 1 은 그룹 기반 MIV 부호화 기법의 BD-rate 를 나타내며, 대조군과 실험군은 동일한 개수의 아틀라스를 사용하도록 설정하였다. High, low BD-rate 는 각각 상위, 하위 4 개 QP 및 대역폭에 대한 BD-rate 를 나타낸다. 전반적으로 그룹 기반 MIV 부호화 기법은 고대역폭에서 높은 효율을 보여주었다. 아틀라스 4 개 사용 시 그룹을 2 개로 분할하고 그룹 당 2 개의 아틀라스를 할당하는 기법이 38.27%의 high IV-PSNR BD-rate 를 기록하여 기존 기법 대비 높은 효율을 보여주었다. 반면, 그룹을 4 개로 분할하고 그룹 당 1 개의 아틀라스를 할당하였을 때는 기존 기법 대비 33.80%의 high IV-PSNR BD-rate 손실이 있었다. MIV 부호화는 그룹 별로 진행되고, 한 그룹 내에서는 기존 MIV 부호화와 마찬가지로 아틀라스 내에 포함되지 못하는 패치들은 버려져 가상 시점 합성 시 일부 정보의 손실로 이어질 수 있음을 고려했을 때, 그룹 기반 MIV 부호화 시에는 그룹을 많이 분할하는 것 보다는 그룹 내 아틀라스 개수를 늘리는 것이 효율적임을 확인하였다. 마찬가지로 아틀라스 6 개 사용 시 6 개의 그룹을 사용하였을 때도 그룹 별 단일 아틀라스 내에 들어갈 수 있는 패치 수가 제한되어 8.68%의 high IV-PSNR BD-rate 손실을 기록하였다. 반면 아틀라스 8 개 사용 시 4 개의 그룹을 사용하고 각 그룹 당 2 개의

아틀라스를 할당하였을 때 기존 기법 대비 7.52%의 high IV-PSNR BD-rate 를 기록하여 그룹 기반 MIV 부호화 기법의 효율을 확인하였다. 향후 그룹 기반 MIV에 선택적 스트리밍 적용 시, 일부 그룹만 전송하여도 가상 시점 합성이 가능하므로 상기 효율의 향상을 기대할 수 있다. 그림 3 은 그룹 적용/미적용 시 율-왜곡 곡선 (rate-distortion curves, RD-curves) 을 나타낸다. 그룹 미적용 시 아틀라스 4 개 사용 대비 6 개, 8 개 사용 시 율-왜곡 효율이 증가했으며, 6 개 및 8 개 사용 시 효율은 비슷하므로 복호기 인스턴스를 줄이기 위해 6 개의 아틀라스가 사용될 수 있다. 그룹 적용 시에는 아틀라스 6 개 대비 8 개 사용 시 율-왜곡 효율이 증가하였으나, 증가폭이 크지 않고 역시 복호기 인스턴스 개수가 증가하므로 복호기 복잡도 역시 고려되어야 한다.

4. 결론

본 논문은 그룹 기반 MIV 부호화 기법을 최신 버전의 TMIV 에서 구현하고 성능 평가를 진행하였다. 대형 실사 컨텐츠에 대한 효율은 아틀라스 6 개 사용 시 수렴하는 경향을 보였고, 그룹을 많이 분할하기보다는 그룹 당 아틀라스 개수를 늘리는 것이 BD-rate 및 율-왜곡 측면에서 효율적임이 확인되었다. 향후 연구에서는 그룹 기반 MIV 에 대한 선택적 스트리밍 기법을 구현 및 실험할 계획이다.

Acknowledgement

이 논문은 과학기술정보통신부에서 시행한 한국전자통신연구원의 연구개발지원사업의 지원을 받아 수행된 연구임 (No.2022-0-00022-001, 초실감 메타버스 서비스를 위한 실사기반 입체영상 공간컴퓨팅 기술 개발).

참고문헌

[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.

[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.

[4] 이상호, 신홍창, 이광순 서정일. 2022. 실시간 재생을 위한 TMIV 디코더의 GPU 구현. 2022 년 한국방송·미디어공학회 하계학술대회, pp. 93-96.

[5] B. Salahieh, S. Bhatia, J. Boyce. 2019. Grouping Implementation in TMIV. Standard ISO/IEC JTC1/SC29/WG11, MPEG/m49859.

[6] S. Lee, J. -B. Jeong, E. -S. Ryu. 2022. Group-Based Adaptive Rendering System for 6DoF Immersive Video Streaming. IEEE Access, vol. 10, pp. 102691-102700.

[7] S. Lee, J. -B. Jeong, E. -S. Ryu. 2022. Efficient Group-Based Packing Strategy for 6DoF Immersive Video Streaming. 2022 International Conference on Information Networking (ICOIN), pp. 310-314.

[8] B. Salahieh, J. Jung, A. Dziembowski. 2021. Test Model 11 for MPEG Immersive Video. Standard ISO/IEC JTC1/SC29/WG4, MPEG/n00142.

[9] V. K. M. Vadakital. 2019. Description of Immersive Video Core Experiments 1. Standard ISO/IEC JTC1/SC29/WG11, MPEG/n18465.

[10] B. Salahieh, S. Bhatia, J. Boyce. 2019. Group-based TMIV. Standard ISO/IEC JTC1/SC29/WG11, MPEG/m49406.

[11] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. DWS-BEAM: Decoder-Wise Subpicture Bitstream Extracting and Merging for MPEG Immersive Video. 2021 International Conference on Visual Communications and Image Processing (VCIP), pp. 1-5.

[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.

[13] 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, MPEG/m58500.

[14] J. Jung, B. Kroon. 2022. Common Test Conditions for Immersive Video. Standard ISO/IEC JTC1/SC29/WG4, MPEG/n00232.

[15] 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 (TCSVT), vol. 32, no. 11, pp. 7575-7591.