

6 자유도 몰입형 영상을 위한 TMIV Geometry Packing 구현

정종범, 이순빈, 최이현, 류은석

성균관대학교 컴퓨터교육학과

{uof4949, soonbinlee, ceh9141, esryu}@skku.edu

Implementing Geometry Packing in TMIV for Six Degrees of Freedom Immersive Video

Jong-Beom Jeong, Soonbin Lee, YiHyun Choi, Eun-Seok Ryu

Department of Computer Science Education, Sungkyunkwan University

요 약

실사 영상 기반의 메타버스 환경을 구축하기 위한 다수의 카메라를 통한 영상 취득 및 부호화, 전송 기술이 활발히 연구되고 있고, 이를 위해 영상 압축 표준화 단체인 moving picture experts group (MPEG) 에서는 MPEG immersive video (MIV) 표준을 개발하였다. 하지만, 현재 널리 사용되는 가상 현실 영상을 스트리밍 가능한 장비의 연산 능력으로는 MIV 기반 몰입형 영상을 스트리밍 시 복호기 동기화 문제가 발생할 수 있다. 따라서 본 논문은 저사양 및 고사양 장비에서 적응적으로 복호기 개수를 조절 가능한 geometry packing 기법을 MIV 의 참조 소프트웨어인 test model for immersive video (TMIV)에서 구현한다. 제안하는 패킹 기법은 지오메트리 영상을 패킹하여 텍스처 영상과 같은 높이를 가지도록 한 후 각각 단일 서브픽처 (subpicture) 로 부호화한다. 이후 부호화된 서브픽처들에 적응적으로 비트스트림 병합이 적용되어 장비의 복호기 사양에 대응한다.

1. 서론

가상 현실 및 증강 현실 기술의 발전에 힘입어, 실사 및 컴퓨터 그래픽 영상을 가상 환경에 구현하고자 하는 메타버스 산업이 주목받고 있다. 기존의 2-D 디스플레이를 포함하여 머리장착형 영상장치인 head-mounted display (HMD) 에서의 실시예에 대한 요구가 증가하고 있고, 이에 초고화질의 몰입형 영상을 제공할 수 있는 효율적인 영상 스트리밍 기술의 중요성이 강조되고 있다. HMD 기반 몰입형 영상 스트리밍 실시예는 기존 2-D 디스플레이 대비 높은 사양을 요구하는데, 가상 현실 환경에서 사용자의 불편감을 최소화하려면 12K 이상의 몰입형 영상을 90 frames per second (FPS) 및 20ms 이내의 지연 시간에 맞춰 제공해야 한다[1]. 제한된 연산 조건에서 상기 요구 사항을 극복하기 위해 영상 부호화 시 압축 수준에서 영상을 직사각형 형태의 타일 (tile) 로 분할하여 사용자 시점 영역 중심으로 스트리밍하는 영상 처리 연구들이 제안되었다[2-6]. 또한 기존에 널리 상용화된 단일 영상만을 재생하여 제한된 자유도를 제공하는 기술을 넘어 사용자에게 움직임 시차 (motion parallax)를 부여하고 몰입감을 극대화하려면 여러 시점에서 촬영된 영상을 동시에 스트리밍 해야 한다. 영상 압축 표준화

단체인 moving picture experts group (MPEG) 에서는 사용자에게 3 축 회전 및 3 축 움직임을 제공하는 상기 실시예를 6 자유도 (six degrees of freedom, 6DoF) 로 정의하고, 6DoF 를 제공하기 위해 픽셀 레이트 (pixel rate)를 감소시키는 연구[7-8], 영상 간 중복성을 픽셀 단에서 제거하는 연구 등이 제안되었다[9].

MPEG-Immersive (MPEG-I) 표준화 그룹은 high-efficiency video coding (HEVC)[10], versatile video coding (VVC)[11] 등의 기존 영상 압축 도구를 사용하면서 다수의 영상으로 이루어진 몰입형 영상의 압축률 향상을 위해 MPEG immersive video (MIV) 표준화를 시작하였고, 현재 버전 2 가 표준화 진행중에 있다[12]. MIV 표준의 참조 소프트웨어인 test model for immersive video (TMIV)[13]는 픽셀 영역에서 영상 간 중복성을 제거하고 잔차 영상을 추출 및 병합하여 아틀라스 (atlas) 영상을 생성하고, 이를 통해 대역폭과 픽셀 레이트를 동시에 절감할 수 있다. MIV 표준은 고사양 및 저사양 장비에서의 원활한 구동을 위해 아틀라스의 최대 개수를 4 개로 제한하였는데, 단일 하드웨어 복호기를 내장한 저사양 장비에서는 4 개 복호기 구동이 어렵다[14]. 따라서, 저사양 장비에서의 MIV 구동을 위해 복호기 개수를 줄일 수 있는 기술이 필요하다.

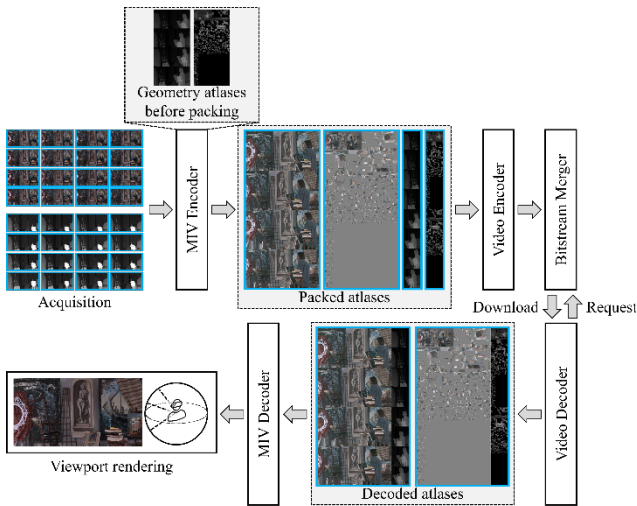


그림 1. 제안하는 TMIV geometry packing 구조도

본 논문은 MIV 기반의 6DoF 몰입형 영상 전송 시 복호기 인스턴스 개수 제어를 위한 geometry packing 기법을 제안한다. 그림 1은 제안하는 TMIV geometry packing 구조도를 나타낸다. 다수의 카메라로부터 취득된, 색상 정보를 나타내는 텍스처 (texture) 와 거리 정보를 나타내는 지오메트리 (geometry) 로 구성된 몰입형 영상은 MIV 부호기에 의해 영상 간 중복성이 제거되고, 아틀라스가 출력된다. MIV 표준 실험 조건에 따라 그림 1의 실시예에서는 지오메트리 아틀라스가 텍스처 아틀라스 대비 $\frac{1}{2}$ 로 다운샘플링되며, 좌측 상단에는 패킹 전 지오메트리 아틀라스가 도시되어 있다. 이후 MIV 부호기에 의해 지오메트리 아틀라스는 픽셀 영역에서 수직 방향으로 분할되어 텍스처 아틀라스와 같은 높이를 가지게 패킹된다. 패킹된 4개의 아틀라스들은 2-D 영상 부호기에 의해 타일 또는 서브픽처 (subpicture) 단위로 부호화된다. 비트스트림 병합기는 부호화된 비트스트림들을 병합하여 2개의 비트스트림을 생성하고, 2-D 영상 부호기에 의해 2개의 부호화된 아틀라스 형태로 출력된다. 이후 MIV 복호기는 부호화된 아틀라스들을 이용하여 뷰 합성 기술을 기반으로 사용자 시점에 대응하는 가상 시점 영상을 생성하고 디스플레이에 도시된다.

본 논문의 구성은 다음과 같다. 2절에서는 MIV의 배경 및 관련 연구로 MIV에서의 프레임 패킹 기법들을 소개한다. 3절에서는 제안하는 geometry packing 기법 설계 및 구현 내용, 결과를 설명한다. 마지막으로 4절에서는 본 논문의 결론을 맺는다.

2. 배경 및 관련 연구

MIV 이전에도 multiview-HEVC (MV-HEVC)[15]와 같이 기존 영상 압축 표준의 확장 형태로 시점 간 압축을 수행하는 기술이 있었으나, 기존 영상 압축 표준 하드웨어 복호기 및 가속 기술을 그대로 사용할 수 없어 널리 사용되기 어렵다. 따라서 MIV는 영상 압축 표준과 독립적으로 동작하도록 설계되었고,

몰입형 영상을 영상 압축 도구로 부호화 및 복호화하기 적합하도록 전처리 및 후처리를 진행할 수 있다.

표 1. MIV 아틀라스 픽셀 레이트 제약조건

Specification	Max. luma sample rate	Max. decoder instances
Low pixel rate (Main10@Level 5.2)	1,069,547,520 (32MP@30FPS)	4
High pixel rate (Main10@Level 6.2)	4,278,190,080 (128MP@30FPS)	4

표 1은 MIV 표준에 따라 생성한 아틀라스의 픽셀 레이트 제약조건을 나타낸다. 현재 MIV의 공통 실험 조건 (common test conditions, CTC) [16]은 표 1의 low pixel rate 제약 조건을 기준으로 하고, 상기 제약조건은 4K 해상도의 영상을 생성할 수 있으며 최대 4개의 복호기 인스턴스를 사용 가능하다. 그러나 복호기 인스턴스를 다수 운용하는 것은 하드웨어 복호기를 하나만 내장한 저사양 장비는 물론이고 고사양 장비에서도 복호기 동기화 문제가 발생할 수 있다. 따라서, MPEG-I에서는 프레임 패킹 (frame packing)을 통해 복호기 인스턴스 수를 줄이는 기법들이 다수 발표되었다.

기존에 발표되고 채택된 프레임 패킹 기법들은 텍스처 아틀라스 하단에 지오메트리 아틀라스를 패킹하는 방식으로 복호기 인스턴스 수를 4개에서 2개로 감소시켰다. 그러나, 서브픽처 기반 병합 방식을 주장한 [17]은 다운샘플링된 지오메트리 아틀라스를 그대로 패킹하여 병합된 픽처 내에 빈 공간을 생성하여 픽셀 레이트를 증가시키고, 서브픽처 영역에 포함되지 않는 빈 공간 때문에 일부 복호기 구현체에서는 복호화 에러가 발생하는 문제가 존재하였다. 픽셀 기반 병합 방식을 주장한 [18]은 지오메트리 아틀라스를 수평 방향으로 분할하여 병합된 픽처 내에 빈 공간을 생성하지 않는다. 그러나, MIV 표준은 텍스처와 지오메트리 간 서로 다른 양자화 매개변수 (quantization parameter, QP)를 사용할 것을 권장하는데, 현재 같은 픽처 내에서 coding tree unit (CTU) 단위로 QP를 조절 가능한 VVC 구현체가 없고, 서브픽처를 사용하더라도 각 서브픽처 크기가 CTU의 배수 (e.g. 128 in VVC)로 나누어 떨어지지 않는 문제로 비트스트림 추출 및 병합 시 오류가 발생하여 사용 가능한 영상 압축 도구의 범위가 제한된다.

3. TMIV Geometry Packing 구현

본 절은 TMIV 소프트웨어에서 구현한 geometry packing 기법의 설계 및 동작 기법에 대해 기술한다. 제안하는 geometry packing 기법은 TMIV v11.0 버전 상에서 구현되었다.

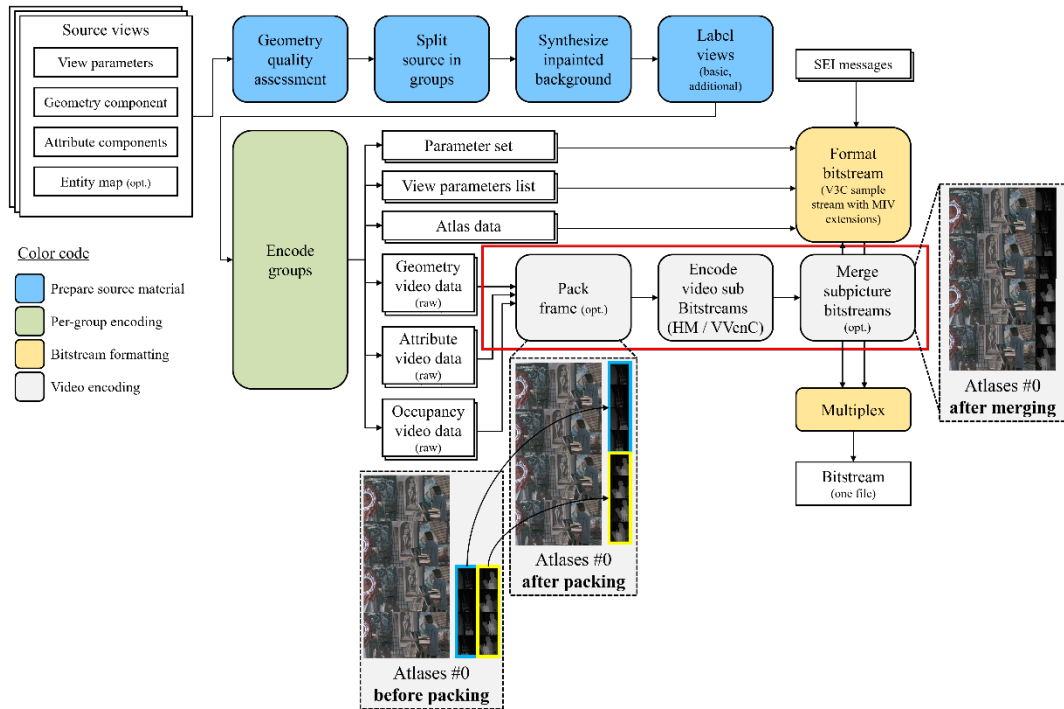


그림 2. Geometry packing 이 적용된 TMIV 부호기 상위 다이어그램

그림 2는 제안하는 geometry packing 기법이 적용된 TMIV 부호기의 상위 다이어그램을 도시한다. 빨간색 실선으로 표시된 영역은 본 논문에서 추가된 처리 요소를 나타낸다. 다수의 카메라로부터 취득된 몰입형 영상은 텍스처 혹은 attribute라 불리는 색상 정보와 지오메트리 혹은 depth라 불리는 거리 정보, 그리고 각 시점의 정보를 담은 카메라 매개변수로 구성되며, 영상 내 객체 정보를 담은 entity map이 포함될 수 있다. 몰입형 영상은 그룹으로 분할될 수 있으며, MIV 부호기에 의해 중복성이 제거되고 잔차 영역이 추출되어 아틀라스의 형태로 출력된다. 이후 지오메트리 아틀라스를 텍스처 아틀라스와 병합하기 적합하게 수직 방향으로 region 단위로 분할하며, 각 region은 그림 2의 좌측 하단에 파란색, 노란색 박스로 도시되었다. Region은 수직 방향으로 정렬되어 픽셀 영역에서 패킹되며 패킹된 지오메트리 아틀라스는 텍스처 아틀라스와 동일한 높이를 가진다. 패킹된 아틀라스들은 HEVC test model (HM) 또는 versatile video encoder (VVenC)를 따르는 4개의 2-D 영상 부호기를 통해 하나의 타일 또는 서브픽처로 부호화한다. 이후 부호화된 서브픽처 비트스트림들을 병합 시 그림 2의 우측 중앙에 도시된 예와 같이 지오메트리 아틀라스는 텍스처 아틀라스의 우측에 패킹된다. 아틀라스의 크기는 표 1에 도시된 픽셀 레이트 내에서 정해지되 너비는 입력된 몰입형 영상의 것과 같고 높이가 조절되는 식으로 정해진다. 따라서 아틀라스의 높이는 CTU 크기의 배수가 아닐 수 있으나 통상 몰입형 영상의 너비는 CTU 크기의 배수로 정해지므로 제안하는 방법은 서브픽처 사용 시 픽처 크기로 인한 에러를 우회 가능하다. HEVC 사용 시 [19]의 타일 병합기를, VVC 사용 시 [20]의

서브픽처 병합기를 통해 본 실시예를 구현할 수 있다. 또한 각 아틀라스를 복원 시 필요한 메타데이터는 visual volumetric video coding (V3C) 비트스트림의 형태로 저장되는데, 아틀라스 서브픽처 비트스트림 병합 시 각 region의 위치가 변경되므로 V3C 비트스트림 수정 역시 진행된다. 이후 병합된 비트스트림과 V3C 비트스트림에 대해 멀티플렉싱 (multiplexing)을 진행하여 단일 비트스트림 생성 후 전송한다. 복호기 단계에서는 2개의 2-D 영상 부호기를 통해 병합된 아틀라스 픽처 출력 후 V3C 비트스트림 정보를 파싱 (parsing) 하여 프레임 언패킹 (frame unpacking)을 진행한다. 이후 복원된 아틀라스 픽처를 이용하여 사용자 시점에 부합하는 가상 시점 영상을 생성한다.

제안하는 기법은 HEVC, VVC 등의 영상 압축 표준과 호환성을 확보하면서 TMIV 상에서 구현될 수 있고, 서브픽처 기반으로 처리하므로 유연한 시스템 설계가 가능하다.

4. 결론

본 논문은 복호기 인스턴스 개수를 절감하기 위한 TMIV에서의 geometry packing 기법을 제안하고 구현하였다. 제안하는 패킹 기법은 지오메트리 아틀라스를 수직으로 분할 및 패킹하고 서브픽처를 이용하여 부호화 및 병합하여 복호기 인스턴스 개수를 절감한다. 제안하는 기법에 의해 부호화된 아틀라스 비트스트림은 TMIV 복호기에 의해 문제없이 복호화됨을 확인하였다. 향후 연구에서는 서브픽처 병합기를 TMIV 내에 통합 구현할 계획이다.

Acknowledgement

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00920-003.(세부 2)중대형 공간용 초고해상도 비정형 플렌옵틱 영상 저장/압축/전송 기술 개발).

참고문헌

- [1] M. -L. Champel, T. Stockhammer, T. Fautier, E. Thomas, R. Koenen. 2016. Quality Requirements for VR. 116th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 116/m39532.
- [2] J. -W. Son, D. Jang, E. -S. Ryu. 2018. 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, pp. 61-66.
- [3] S. Lee, D. Jang, J. B. Jeong, E.-S. Ryu. 2019. Motion-constrained tile set based 360-degree video streaming using saliency map prediction. In Proceedings of the 29th ACM Network and Operating System Support for Digital Audio and Video, pp. 20-24.
- [4] J. -B. Jeong, S. Lee, I. -W. Ryu, T. T. Le, E. -S. Ryu. 2020. Towards Viewport-dependent 6DoF 360 Video Tiled Streaming for Virtual Reality Systems. In Proceedings of the 28th ACM International Conference on Multimedia, pp. 3687-3695.
- [5] J. -B. Jeong, S. Lee, I. Kim, S. Lee, E. -S. Ryu. 2020. Implementing VVC Tile Extractor for 360-degree Video Streaming Using Motion-Constrained Tile Set. Journal of Broadcast Engineering, vol. 25, no. 7, pp. 1073-1080.
- [6] J. -B. Jeong, S. Lee, I. Kim, E. -S. Ryu. 2021. Implementing Viewport Tile Extractor for Viewport-Adaptive 360-Degree Video Tiled Streaming. In Proceedings of the 2021 International Conference on Information Networking, pp. 8-12.
- [7] J. B. Jeong, D. Jang, J. Son, E. -S. Ryu. 2018. 3DoF+ 360 Video Location-Based Asymmetric Down-Sampling for View Synthesis to Immersive VR Video Streaming. Sensors, vol. 18, no. 9, pp. 3148.
- [8] D. Mieloch, P. Garus, M. Milovanović, J. Jung, J. Y. Jeong, S. L. Ravi, B. Salahieh. 2022. Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video. IEEE Transactions on Circuits and Systems for Video Technology.
- [9] J. -B. Jeong, S. Lee, D. Jang, E. -S. Ryu. 2019. Towards 3DoF+ 360 Video Streaming System for Immersive Media. IEEE Access, vol. 7, pp. 136399-136408.
- [10] 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, 22(12), pp. 1649-1668.
- [11] 이종석, 박준택, 최한솔, 변주형, 심동규. 2019. Overview of VVC. 방송과 미디어, 24 권 4 호, pp. 10-25.
- [12] 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.
- [13] B. Salahieh, J. Jung, A. Dziembowski. 2021. Test Model 11 for MPEG Immersive Video. Standard ISO/IEC JTC1/SC29/WG4, MPEG/n00142.
- [14] 2019. Thoughts on Immersive Media Decoding Interface for VVC. Standard ISO/IEC JTC1/SC29/WG11, MPEG/n18438.
- [15] G. Tech, K. Müller, J. -R. Ohm, A. Vetro, Y. -K. Wang. 2016. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49.
- [16] J. Jung, B. Kroon. 2022. Common Test Conditions for Immersive Video. Standard ISO/IEC JTC1/SC29/WG4, MPEG/n00169.
- [17] L. Kondrad, V. K. M. Vadakital, L. Ilola. 2020. CE-1.3: frame packed video sub-bitstream type. Standard ISO/IEC JTC1/SC29/WG11, MPEG/m54274.
- [18] B. Salahieh, G. Naf, J. Boyce. 2021. Frame Packing Implementation in TMIV. Standard ISO/IEC JTC1/SC29/WG4, MPEG/m56827.
- [19] HEVC Tile Merger. 2022. Retrieved from <https://gitlab.com/MCSLJeong/hevctilemerger>
- [20] J. -B. Jeong, S. Lee, E. -S. Ryu. 2022. [MIV] Geometry Packing Implementation in TMIV for Frame Packed Video. Standard ISO/IEC JTC1/SC29/WG4, MPEG/m59442.