

몰입형 비디오 부호화를 위한 패치 패킹 정보의 효율적인 표현

임성균, 윤용욱, 김재곤

한국항공대학교

sglim@kau.kr, yuyoon@kau.kr, jgkim@kau.ac.kr

Efficient Representation of Patch Packing Information for Immersive Video Coding

Sung-Gyun Lim, Yong-Uk Yoon, and Jae-Gon Kim

Korea Aerospace University

요 약

MPEG(Moving Picture Experts Group) 비디오 그룹은 사용자에게 움직임 시차(motion parallax)를 제공하면서 3D 공간 내에서 임의의 위치와 방향의 시점(view)을 렌더링(rendering) 가능하게 하는 6DoF(Degree of Freedom)의 몰입형 비디오 부호화 표준인 MIV(MPEG Immersive Video) 표준화를 진행하고 있다. MIV 표준화 과정에서 참조 SW 인 TMIV(Test Model for Immersive Video)도 함께 개발하고 있으며 점진적으로 부호화 성능을 개선하고 있다. TMIV 는 여러 뷰로 구성된 방대한 크기의 6DoF 비디오를 압축하기 위하여 입력되는 뷰 비디오들 간의 중복성을 제거하고 남은 영역들은 각각 개별적인 패치(patch)로 만든 후 아틀라스에 패킹(packing)하여 부호화되는 화소수를 줄인다. 이때 아틀라스 비디오에 패킹된 패치들의 위치 정보를 메타데이터로 압축 비트열과 함께 전송하게 되며, 본 논문에서는 이러한 패킹 정보를 보다 효율적으로 표현하기 위한 방법을 제안한다. 제안방법은 기존 TMIV10.0 에 비해 약 10%의 메타데이터를 감소시키고 종단간 BD-rate 성능을 0.1% 향상시킨다.

1. 서론

최근 사용자에게 시점의 자유도를 제공하는 몰입형 비디오는 몰입감 있는 시각적 경험을 제공하는 새로운 미디어 유형으로 각광받고 있다. MPEG 에서는 6DoF(Degree of Freedom)까지의 시점 자유도를 제공하는 것을 목표로 몰입형 비디오를 부호화하기 위한 MPEG Immersive Video(MIV) 표준화 작업을 활발히 진행하고 있다[1], [2]. 3DoF 의 자유도를 제공하는 360 도 비디오와 달리 6DoF 비디오는 사용자의 시점 변화에 따른 뷰포트(viewport)를 제공하는 움직임 시차(motion parallax)를 지원한다. 움직임 시차 기능을 지원하기 위해 3D 공간에서 시점 위치 변화에 따라 모든 시점에서 가상의 시점을 합성할 수

있어야 하고, 이를 위해 6DoF 비디오 렌더링에는 깊이(depth) 정보를 기반으로 임의 시점의 비디오를 합성한다. 따라서 6DoF 몰입형 비디오는 3D 공간 내에서 동시에 여러 시점에서 촬영된 비디오와 그에 상응하는 깊이 비디오가 필요하다. 이렇게 몰입형 비디오는 다중 시점의 비디오로 구성되어 매우 큰 데이터 크기를 가지게 되어 보다 효율적인 압축 기법이 필수적으로 요구된다.

MPEG 비디오 그룹에서는 대용량의 6DoF 비디오를 효율적으로 압축하여 전송하기 위해서 다양한 부호화 방법이 연구되고 있고, 최근 DIS(Draft IS)[3]와 참조 SW 인 TMIV(Test Model for Immersive Video)10.0 버전[4]을 발간했다. TMIV 의 몰입형 비디오 압축을 위한 주요 개념은 입력되는 시점 비디오들 간의 상관관계가 높은 특성을 이용하여 압축 전단계에서 모든 입력 시점 비디오들간의 중복성을 제거하여 부호화 할 화소수를

줄이는 것이다. 데이터의 총량을 줄이는 것이다. 즉, 입력되는 뷰 비디오들 간의 중복성을 제거하고 남은 영역들을 각각 패치(patch)로 만든 후 아틀라스에 패킹(packing) 한다. 이때 아틀라스 비디오에 패킹된 패치들의 위치 정보를 메타데이터로 압축 비트열과 함께 전송하게 되며, 본 논문에서는 TMIV 에서 이러한 패킹 정보를 보다 효율적으로 표현하기 위한 방법을 제안한다.

2. Test Model for Immersive Video

그림 1 은 TMIV 부호화기의 전체적인 구조를 보여준다. 먼저 시점 라벨링(view labeling) 과정에서 서로 다른 위치의 입력 시점 비디오들을 몇 개의 그룹으로 분류하고 각 그룹 내의 뷰들을 기본 시점(basic view)과 추가 시점(additional view)으로 구분하고, 각 그룹별로 기본 시점과 중복성이 있는 추가 시점 비디오들의 영역들이 제거되는 프루닝(pruning) 과정이 수행된다. 프루닝 과정 후에 남은 영역들은 직사각형 형태의 개별적인 패치로 생성된다. 이 때, 하나의 패치에는 전송되어야 할 잔여 픽셀들을 포함하는 유효 영역(valid region)과 프루닝에 의해 제거된 부분인 비유효 영역(invalid region)이 모두 포함된다.

프루닝 과정에서 추가 시점으로부터 생성된 패치는 기본 시점 비디오와 함께 소수의 아틀라스로 패킹된다. 패킹 되는 패치들은 유효영역이 서로 겹치지 않는 한에서 최대한 조밀하게 패킹되어 패치가 차지하는 공간을 최소화한다. 또한, 패치가 차지하는 공간을 최대한 줄이기 위해 패치들은 회전된 후 패킹될 수 있으며, 이때 각 패치가 패킹된 위치 및 회전 정보는 메타데이터로 구성되어 부호화된 아틀라스의 비트스트림과 같이 전송된다. 이 메타데이터 정보는 TMIV 복호화기가 아틀라스에 패킹된 패치들을 추가 시점 비디오의 원래 위치로 복원할 수 있도록 해준다.

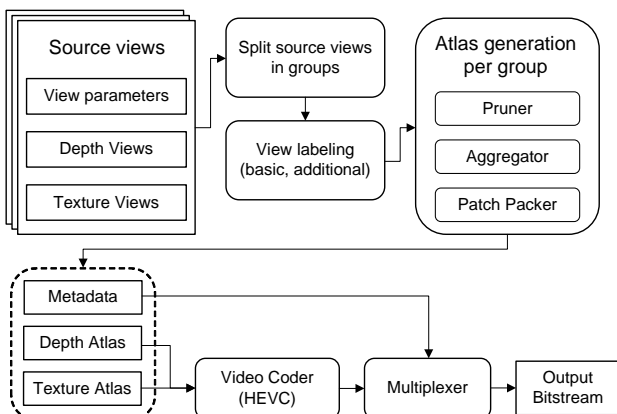


그림 1. TMIV 부호화기의 전체 구조도[4]

3. 제안기법

TMIV 복호화기로 전송되는 메타데이터 중에서 각각의 패치가 아틀라스에 패킹된 위치 정보는 각 직사각형 패치의 왼쪽 상단 정점의 좌표들로 표현된다. 그리고 TMIV 에서 각 패치의 패킹 위치 정보는 지수-골롬 부호화(Exp-Golomb coding) 방식을 통해 이진 부호화 된다. 지수-골롬 부호화는 절대값이 작은 값들을 부호화 할 때 더 효율적인 방식이기 때문에 패치 패킹 위치 정보들을 패치들 간의 상대적인 위치 정보로 표현하여 부호화 하는 것이 효율적이다. 본 논문에서는 패치가 패킹된 위치를 보다 효율적으로 표현하여 부호화 하고 메타데이터의 양을 감소시키는 방법을 제안한다[5].

그림 2 는 임의의 2 개의 패치가 한 아틀라스에 패킹되는 경우의 예시를 보여준다. 그림에서 (x_n, y_n) 는 n 번째 패치, (x_{n+1}, y_{n+1}) 는 $(n+1)$ 번째 패치가 아틀라스에 패킹된 위치를 나타낸다. 두 패치의 패킹 위치 정보들이 부호화되어 복호화기로 전송될 때, 패치들이 원점을 기준으로 위치한 좌표들로 표현된다. 그런데 $(n+1)$ 번째 패치의 위치는 n 번째 패치의 위치와의 차이값인 $(\Delta x, \Delta y)$ 으로 표현하여 전송할 수 있다. 복호화기에서는 위치 간의 차이값을 가지고 n 번째 패치의 위치를 참조하여 $(n+1)$ 번째 패치의 위치를 복원할 수 있다. 패치들은 크기가 큰 것부터 작은 것 순서대로 아틀라스에 최대한 조밀하게 패킹 되기 때문에 패킹되는 순서 상 인접한 패치 간의 거리는 가까울 확률이 높다. 따라서, 상대적인 위치 정보로 표현하는 것이 더 절대값이 작은 값들로 표현할 수 있다.

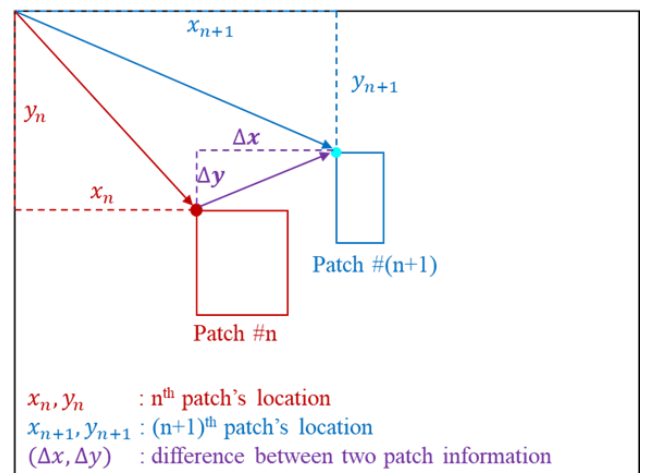


그림 2. 아틀라스에 두 개의 패치가 패킹되는 경우의 예

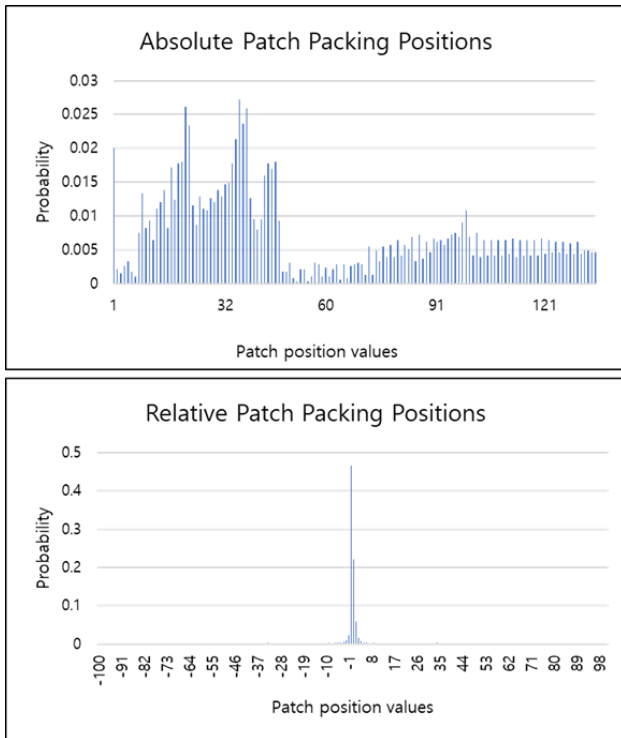


그림 3. ‘Painter’ 시퀀스의 각 패치의 패킹 위치 값 분포 (위: 기존방법, 아래: 제안방법)

그림 3 은 ‘Painter’ 시퀀스의 아틀라스를 생성할 때, 아틀라스에 패킹된 패치의 위치 좌표값들의 원본값과 패킹된 순서상 인접한 패치 사이의 차이값의 분포 차이를 보여준다. 패치 패킹 위치의 원본값의 분포는 전체 범위에 걸쳐 비교적 고른 분포를 보이는 반면, 패킹 순서상 인접한 패치 간의 위치 좌표 차이값의 분포는 0 에 가까운 영역에 집중적으로 분포된다. TMIV 에서 사용하는 지수-골롬 부호화 방식은 부호 구별을 위한 비트를 더 사용하더라도 절대값이 작은 값들로 표현된 경우에 더 효율적인 작용하기 때문에 제안 방법을 통해 메타데이터의 양을 크게 감소시킬 수 있다.

4. 실험결과

표 1 은 제안된 방법으로 생성한 메타데이터의 양이 기존 TMIV 에 의해 생성된 메타데이터의 양과 비교한 결과이다. 제안방법을 TMIV10.0 에 구현하여 MIV CTC(Common Test Conditions) [6]에 따라 성능확인 실험을 수행하였다. 제안방법은 모든 시퀀스에서 메타데이터의 양을 상당히 감소하였으며 평균적으로 9.8% 메타데이터 절감을 얻었다.

표 1. 기존방법 대비 제안방법에 의한 메타데이터 감소량

Sequence		Anchor (Kb)	Proposed (Kb)	Reduction ratio
ClassroomVideo	A	173.4	153.9	-11.23 %
Museum	B	64.5	57.9	-10.29 %
Fan	O	114.0	102.3	-10.26 %
Kitchen	J	164.2	147.9	-9.92 %
Painter	D	104.9	92.5	-11.82 %
Frog	E	97.9	88.9	-9.23 %
Chess	N	94.9	89.0	-6.22 %
Group	R	140.5	127.3	-9.39 %
Average				-9.80 %

4. 결론

본 논문에서는 몰입형 비디오 부호화 효율 개선을 위하여 효율적인 패치 패킹 위치 정보 표현 기법을 제시하였다. 제안 기법은 패치의 패킹 위치 정보들을 패치들 간의 상대적인 위치 좌표로 표현함으로써 효율적이 부호화를 가능하게 했다. 실험결과에서 제안기법은 메타데이터량을 9.8% 감소시킨 것을 확인할 수 있었다.

Acknowledgement

이 논문은 2021 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2020R1A6A13073358)

References

- [1] “MPEG-I Use Cases for Omnidirectional 6DoF, Windowed 6DoF, and 6DoF,” ISO/IEC JTC1/SC29/WG11, N16768, Apr. 2017.
- [2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, “Standardization Status of Immersive Video Coding,” IEEE J. Emerg. Select. Topics Circuits Syst., vol. 9, no. 1, pp. 5-17, Mar. 2019.
- [3] “Text of ISO/IEC DIS 23090-12 MPEG Immersive Video,” ISO/IEC JTC1/SC29/WG04, N00049, Jan. 2021.
- [4] B. Salahieh, J. Jung, and A. Dziembowski (Eds.), “Test Model 10 for Immersive Video,” ISO/IEC JTC1/SC29/WG04, N0112, Jul. 2021.
- [5] S. Lim, H. Kim, and J. Kim, “[MIV] CE1-related: Efficient Representation of Patch Packing Information using Differential Position,” ISO/IEC JTC1/SC29/WG04, m56004, Jan. 2021.
- [6] J. Jung, B. Kroon, “Common Test Conditions for MPEG Immersive Video,” ISO/IEC JTC1/SC29/WG04, N00051, Jan. 2021.