

프루닝 클러스터 그래프 구성 전략에 따른 몰입형 비디오 압축 성능 분석

이순빈, 정종범, 류은석

성균관대학교 컴퓨터교육학과

soonbinlee@skku.edu, uof4949@skku.edu, esryu@skku.edu

Efficient Pruning Cluster Graph Strategy for MPEG Immersive Video Compression

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

Department of Computer Science Education, Sungkyunkwan University

요 약

MPEG Immersive Video (MIV) 표준화 기술은 다시점 영상 부호화 시 비디오 코덱의 부담을 최소화하기 위해 각 시점 영상의 차분 정보만을 표현하는 처리 기술을 바탕으로 하고 있다. 본 논문에서는 시점 간 중복성 제거를 진행하는 과정인 프루닝(pruning) 과정에서 복잡도 절감을 위해 병렬처리에 용이하도록 구성되는 프루닝 클러스터 그래프에 대해 서술하고, 각 클러스터 그래프 별 구성 전략에 따른 성능 분석을 진행한다. 클러스터 그래프 내에서 중복성 제거를 진행하지 않고 완전한 정보를 보존하는 바탕 시점(basic view)의 개수가 적게 포함될수록 처리할 전체 픽셀 화소율(pixel rate)은 감소하지만, 복원 화질 역시 감소하며 프루닝 복잡도는 증가하는 경향을 보인다. 실험 결과를 통해 프루닝 클러스터 그래프 구성에 따른 트레이드오프를 탐색하고, 최적화된 그래프 구성 전략에 따라 몰입형 비디오의 효율적인 전송이 가능함을 보인다.

1. 서론

가상현실 (Virtual Reality, VR) 기술 등에 대한 관심이 증대됨에 따라, 보다 몰입할 수 있는 실감형 미디어 압축 기술에 대한 논의가 활발하게 이루어지고 있다. Moving Picture Experts Group (MPEG)에서는 몰입형 미디어를 표준화하기 위해 MPEG-I (Immersive) 서브그룹에서 가상현실 비디오 표준화를 진행중에 있으며, 여러 위치에서 취득한 다시점 영상을 처리하는 MPEG Immersive Video (MIV) 기술에 관한 표준화가 진행되고 있다. 이러한 표준화의 일환으로 MPEG-I 그룹은 6 degrees of freedom (DoF)을 제공하는 몰입형 미디어를 지원하는 테스트 모델(Test Model for Immersive Video, TMIV)를 정의하고, 개발중에 있다[1,2].

TMIV 의 처리 기술은 픽셀 화소율(pixel rate)의 감소를 목적으로 두고 있다. 기존의 MV-HEVC, 3D-HEVC 에서 다수의 디코더 인스턴스가 필요한 부담이 있는 반면에, MIV 에서는 다시점 영상들을 전처리를 통해 픽셀 화소율의 부담을 낮추고자 한다. 기본이 되는 영상 처리 모델은 정보를 가장 많이 담고 있는 시점의 영상을 계산해 바탕 시점(basic view)으로 지정한 후, 나머지 시점의 영상들은 시점 간 중복성 제거 과정을 통해 바탕 시점과의 차분 정보만을 보존하는 영상들로 따로 분리하여 추가 시점(additional view)으로 구분한다[3].

이 때, 시점 간의 차분 정보를 계산하여 다른 정보를 담고 있는 패치들만을 계산하는 과정을 프루닝 (pruning)이라고 한다. 프루닝을 통해 분리된 추가 시점의 영상 정보들을 패치(patch)라고 표현하며, 복원 시에 다시 이 패치를 활용해 원래의 전체 영상들을 복원할 수 있도록 위치 정보 등을 메타데이터로 함께 포함하여 전송한다. 몰입형 비디오 기술은 이와 같은 시점 간 중복성 제거 전처리 과정으로 현재 4K 해상도 크기로 제한된 비디오 디코더의 개수를 비약적으로 절감할 수 있는 장점을 가진다. 이 때, 프루닝 과정에서 바탕 시점의 개수 선택과 추가 시점의 순서 등에 따라 압축 효율에 큰 영향을 미칠 수 있다. 그림 1 은 프루닝 그래프의 예시를 나타내며, 바탕 시점은 추가 시점과의 프루닝을 위해 사전에 선택되어 처리된다.

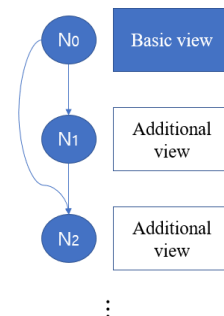


그림 1. 프루닝 그래프 다이어그램

2. 배경 및 관련연구

프루닝 클러스터는 MPEG 131 차 회의에서 바탕 시점 할당(basic view allocation) 기준에 대한 논의와 함께 제안된 개념이다. 이전 회의에서 TMIV 프루닝 기술에 대한 Core Experiments (CE)를 진행한 결과, Pixel rate 제한 조건을 만족하는 선에서 최대한 바탕 시점의 개수가 많을수록 복원 렌더링 화질에서 좋은 성능을 보였음을 제시하였다[4,5]. 이에 따라 TMIV 6 버전에서부터는 바탕 시점 포함 기준에 변화가 일어남과 함께, 높아진 복잡도를 감소시키기 위한 프루닝 그래프를 클러스터 단위로 독립적으로 처리하고자 하는 제안이 채택되어 테스트 모델 소프트웨어에 통합되었다.

```

( v12 v15 ) <- v7 <- v13 <- v14 <- v8 <- v11 <- v9 <- v10
( v0 v3 ) <- v4 <- v2 <- v5 <- v6 <- v1 <- s0
    
```

그림 2. Pruning 그래프 형성 예시 (Painter sequence)

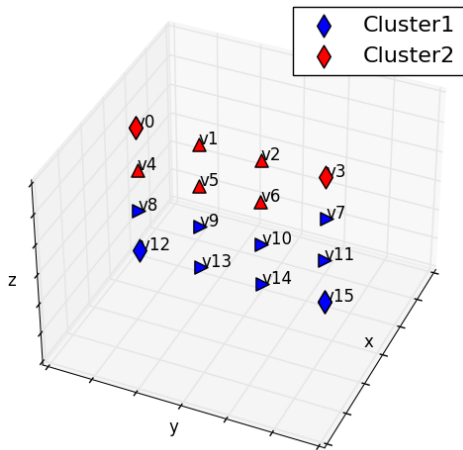


그림 3. 클러스터 카메라 좌표 예시 (Painter sequence)

그림 2 는 *Painter* 시퀀스의 프루닝 클러스터 그래프의 형성 예시를 나타낸다. 각 클러스터의 그래프들은 서로 독립적으로 프루닝 과정을 처리하게 된다. 이에 따라 프루닝 처리 횟수가 감소하여 전체 중복성은 다소 늘어나게 되지만, 병렬 처리 등을 통해 TMIV 인코딩 복잡도를 절감할 수 있게 된다. 괄호가 포함된 시점은 바탕 시점을, 그렇지 않은 시점은 추가 시점을 나타내며 그림 3 에서도 바탕 시점에 대한 구별을 표시하였다. 그림 3 은 앞서 구성된 클러스터의 그래프를 토대로 시각화된 카메라 좌표의 시각화를 나타낸다. 클러스터 구성에 있어서는 바탕 시점과 추가 시점의 선택 방법이 영향을 미친다는 것을 확인할 수 있으며, 보다 효율적인 구성 전략을 탐색하기 위해 TMIV에서의 시점 선택 방법에 대해 간략히 서술한다.

바탕 시점 선택 방법: TMIV 에서 바탕 시점을 할당하기 위해서는 우선 제한된 Pixel Rate 조건과 입력 영상 해상도를

토대로, 바탕 시점의 개수를 계산한다. 이후, 결정된 최적의 바탕 시점의 개수를 토대로 다른 입력 영상들 간의 비용 함수(cost function)를 계산한다. 테스트 모델의 비용 함수는 k-medoids 알고리즘을 기반으로, 바탕 시점 영상과 다른 입력 영상들 간의 적절한 거리를 유지할 수 있도록 바탕 시점이 선택된다. 130 차 회의까지의 테스트 모델 (TMIV 5)에서는 바탕 시점 선택 단계에서부터 입력 영상 간의 overlap score 를 계산해 이를 활용하였으나, TMIV 6 에서부터의 테스트 모델에서는 복잡도 문제로 영상의 거리 정보만을 토대로 비용 함수를 구성한다.

추가 시점 선택 방법: 앞서 서술한 과정에서 바탕 시점으로 선택되지 않은 나머지 입력 영상은 모두 추가 시점으로 간주되며, 프루닝 과정에서 추가 시점의 순서에 따라 압축 효율에 영향을 미치게 된다. 추가 시점은 미리 결정된 바탕 시점에 projection 하여 두 영상의 overlap score 를 계산하고, 해당 점수가 높은 추가 시점부터 순서대로 프루닝 그래프를 이루도록 구성된다. 프루닝 그래프에서는 계층적으로 이전 노드의 모든 바탕 시점과 추가 시점이 다음 노드의 추가 시점과의 프루닝 과정을 거치므로, 그래프의 노드가 많을수록 복잡도가 크게 증가하게 된다.

클러스터 구성 방법: 단순하게 각 클러스터에 포함되는 영상의 개수가 고르게 분배되도록 구성하는 전략이 구현되어 있다. 추가 시점 선택 시에 overlap score 뿐만 아니라, 클러스터 개수를 함께 고려하여 그래프를 구성하도록 함으로써 각 클러스터에 영상이 균등하게 분배되도록 구성된다.

3. 제안 방법

제안하는 방법은 클러스터 내에서 최대 바탕 시점의 개수를 변경한다. 기존의 그룹 기반 방법에서는 전체 영상을 일정 단위 개수의 시점으로 나누어 별도로 처리를 진행하면 깊이 맵이 부정확한 자연 영상(Natural Content)에서는 렌더링 성능의 향상을 보인 실험 결과가 존재한다[6]. 제안하는 방법 역시 클러스터 간 중복성을 높여 최종 결과의 픽셀 화소율(pixel rate)과 전체 필요 비트레이트는 다소 늘어나게 되지만, 렌더링 화질에서의 향상 결과를 보인다. TMIV 설정 파라미터인 'maxBasicViewsPerGraph' 값을 조절하여, 클러스터 내에 포함될 수 있는 최대 바탕 시점의 개수에 따른 변화에 대해 실험을 진행하였으며, TMIV 내에서 아틀라스 생성 시에 적합한 바탕 시점의 개수를 산출하여 제한하기 때문에, 바탕 시점의 비율 역시 파라미터 값을 조절하였다. 제안 방법의 예시로 바탕 시점의 개수를 N=1 으로 제한하였을 때, 각 클러스터는 하나의 바탕 시점만을 포함하면서도 비슷한 개수의 영상을 포함해야하는 조건으로 인해 클러스터의 개수가 늘어나게 된다. 이 특성은 향후 물임형 비디오의 스트리밍 시나리오에서 독립적인 클러스터 전송/복원의 동작을 포함할 수 있다[7].

```

( v15 ) <- v10 <- v7 <- v14 <- v11
( v12 ) <- v9 <- v13 <- v8
( v3 ) <- v5 <- v2 <- v6
( v0 ) <- v4 <- v1 <- s0
    
```

그림 4. 제안 방법의 Pruning 그래프 (Painter sequence, N=1)

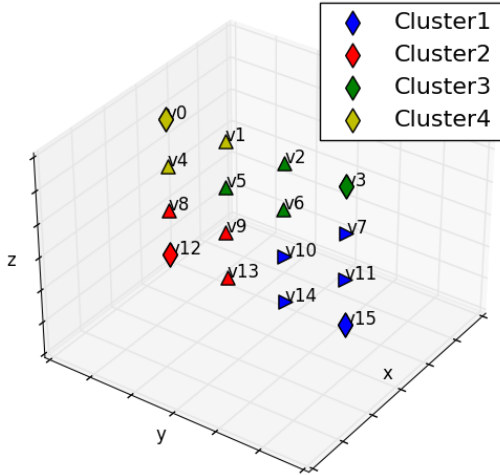


그림 5. 변경된 클러스터 카메라 좌표 예시 (Painter sequence)

4. 실험 결과

실험에 사용한 소프트웨어는 HM 16.20, MIV 8.0 으로 현재 MPEG-I 에서 진행중인 표준 CTC (Common Test Condition)에 정의되어 있는 영상 중 자연 영상 카테고리에 포함되어 있는 Painter, Frog, Carpark, Fencing 에 대하여 실험을 진행하였으며, 17 프레임으로 진행하였다[8]. 양자화 파라미터 (quantization parameter, QP) 로는 CTC 에서 타겟 비트레이트에 따라 정의된 시퀀스 별 QP 를 준수하여 진행하였다. Anchor 의 'maxBasicViewsPerGraph' 값은 N=3 이며, 클러스터 그래프 변화에 따른 성능을 관찰하기 위해 N=1,2 에 대한 분석을 진행하였으며, 전체 TMIV 의 바탕 시점 비율을 조절하는 'maxBasicViewFraction' 값 또한 기존 Anchor 의 0.5 에서부터 조절하며 진행하였다[9].

바탕 시점 비율이 늘어날수록 압축 효율은 향상되는 것으로 관찰되며, 클러스터 개수가 많은 N=1 에서 일부 시퀀스의 경우 Anchor 대비 압축 성능 향상이 있는 것을 관찰할 수 있다. 바탕 시점 비율에 따른 성능 향상 영향 또한 관찰하기 위하여 N=2,0.75 경우에 대해서도 실험을 진행하였으며, 기존 N=1 대비 높은 성능 향상을 보여주지만, Painter 와 Carpark 등 일부 시퀀스에서는 오히려 더 낮은 성능 향상을 나타내었다. 이는 현재 프루닝 그래프의 구성 전략이 압축 효율 측면에서 좀 더 최적화될 여지가 있음을 나타내며, 대부분의 시퀀스에서

독립적으로 처리되는 클러스터 개수가 증가하는 오버헤드 대비 렌더링 화질 향상이 크다는 것을 나타낸다.

Sequence	N=1, 0.5	N=1, 0.75	N=2,0.75
Painter	-5.87%	-15.14%	-14.15%
Frog	8.90%	-4.25%	-6.55%
Carpark	-3.54%	-17.64%	-11.22%
Fencing	-12.25%	-22.29%	-27.54%

표 1. Anchor 대비 제안하는 방법에서의 PSNR BD-rate saving(%),

5. 결론

본 논문은 프루닝 클러스터 그래프 구성에 따른 몰입형 비디오의 압축 성능을 분석하였다. 제안하는 방법을 통해 클러스터 내의 바탕 시점을 제한할 경우, 클러스터의 개수가 늘어나 처리해야할 중복성이 늘어나는 경향을 보인다. 그러나 오버헤드 대비 일부 시퀀스의 경우 화질 향상으로 인해 전체 BD-rate 에서 이득을 얻을 수 있음을 보였다. 각 클러스터는 TMIV 에서 독립적으로 처리되는 단위로 향후 적응적 스트리밍 또는 사용자 시점 기반 스트리밍 연구에 활용할 계획이다.

Acknowledgement

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

참고문헌

- [1] ISO/IEC JTC1/SC29/WG11/N18654, "Requirements for Immersive Media Access and Delivery", MPEG 127, Jul. 2019.
- [2] <https://gitlab.com/mpeg-i-visual/tmiv>.
- [3] ISO/IEC JTC1/SC29/WG11/N19483, "Test Model 8 for MPEG Immersive Video," MPEG 133, Jan. 2021.
- [4] ISO/IEC JTC1/SC29/WG11/M54145, "[MIV] Philips CE-2 related response "basic view allocator", MPEG 131, May. 2020.
- [5] ISO/IEC JTC1/SC29/WG11/M54754, "Basic view allocator with disjoint pruning graph", MPEG 131, May. 2020.

- [6] ISO/IEC JTC1/SC29/WG11/M54151, "Grouping and anchor study on MIV content", MPEG 131, Jun, 2020.
- [7] Soonbin Lee, Jong-Beom Jeong, and Eun-Seok Ryu. (2022). Group-based Adaptive Rendering for 6DoF Immersive Video Streaming. JOURNAL OF BROADCAST ENGINEERING, 27(2), 216-227.
- [8] ISO/IEC JTC1/SC29/WG11/N19484, "Common Test Conditions for MPEG Immersive Video," MPEG 131, Jul. 2020.
- [9] ISO/IEC JTC1/SC29/WG11/N19492, "Exploration Experiments on Processing for Future MPEG Immersive Video", MPEG 131, Jul. 2020.