

이머시브 비디오 테스트 모델에서의 프루닝 기법의 개선

*신홍창 **윤준영 *이광순 *음호민 *서정일

*한국전자통신연구원 **한양대학교

*hcshin@etri.re.kr

Enhancement of Pruning Order Determining for Immersive Video Test Model

*Shin, Hong-Chang **Yun, Junyoung *Lee, Gwangsoon *Eum, Homin *Seo, Jungil

*Electronics and Telecommunications Research Institute **Hanyang University

요약

이머시브 비디오 서비스를 위해 MPEG-I Visual 그룹에서는 3DoF+ 기술과 관련하여 MIV(Metadata for Immersive video)의 표준화를 진행하고 있으며, 이를 위해 다시점 영상 및 전방위 장면을 촬영한 ERP 규격의 영상들이 주어진 경우에 운동시차를 제공할 수 있는 영상 합성 레퍼런스 소프트웨어인 TMIV SW를 제공한다. TMIV는 기본적으로 송신부인 인코더와 수신부인 디코더로 구성이 되어있으며, 인코더에서 가장 중요한 기능은 다수의 입력 시점영상 간의 중복된 데이터를 찾아내서 제거하는 프루닝 과정이다. 프루닝 방법에 따라 데이터 전송량과 디코더에서의 합성 품질이 달라지기 때문에 인코더에서 핵심이라고 할 수 있다. 본 논문은 인코더의 프루닝의 효율을 높이기 위해 전체 흐름도에서 프루닝 순서 변경 과정을 추가하고 그 과정에서 시점 영상 간 중첩 영역을 계산하여 이를 토대로 프루닝 순서를 결정하는 방법을 제안하였고 이를 통해 데이터 압축률이 향상됨을 확인할 수 있었고, 또한 수신부에서 영상 합성의 품질이 달라짐을 확인할 수 있었다.

1. 서론

지난 2016년 10월 MPEG 청두 회의에서 시작한 이머시브 비디오 표준화 논의는 MPEG-I 프로젝트명으로 visual, system 그룹 등에서 현재까지 활발하게 진행중에 있다.[1]. MPEG-I 프로젝트는 이머시브 비디오 서비스를 3DoF, 3DoF+, 6DoF인 세 단계로 구분하여 각각에 대한 타임라인을 기반으로 진행되고 있다. 3DoF는 고정된 중심축을 기준으로 3축 회전운동만을 지원하고, 여기에 약간의 머리 움직임으로 발생하는 제한된 병진 운동을 지원하는 3DoF+, 실생활과 유사한 가상 환경을 구축하여 자유롭게 이동이 가능한 6DoF로 구분된다.

3DoF 비디오 표준화는 2019년부터 본격 시작하여 지난 2019년 3월 제네바 회의에서 CfP 대응을 통해 제안된 기술들을 바탕으로 테스트 모델인 TMIV(Test Model for Immersive Video) 버전 1.0로 통합하였으며, 현재 2020년 6월에는 TMIV 5.0까지 구현되어 MIV라는 이름으로 WD문서 작성을 진행하고 있다.

2. TMIV 3.0 구조

그림 1은 TMIV 3.0 인코더와 디코더를 보인다[2]. 이머시브 비디오 서비스를 위한 비디오 규격의 다중 입력 영상과 그에 대응되는 깊이 정보 영상이 입력되면, 이를 View optimizer 모듈에서 기본 시점(Basic view)과 부가 시점(Additional view)을 선정한다. 기본 시점 영상은 영상의 모든 화소가 전달되며, 부가 시점 영상은 기본 시점과 선순위 부가

시점 영상을 기준으로 중복된 화소를 검출하여 제거하는 프루닝 (Pruning) 과정을 거치게 되어 화소의 양이 줄어들게 된다.

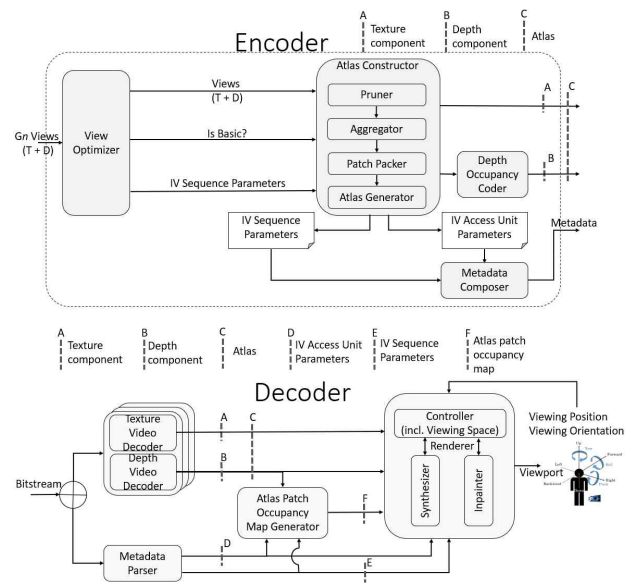


그림 1. TMIV 3.0 구조

프루닝 과정을 거치면 부가 시점 영상들의 화소가 제거되며, 남은 화소들은 Aggregator 모듈에서 intraperiod 단위로 군집화

(clustering) 과정을 거쳐 직사각형의 패치(Patch)가 생성되고 Patch packer 모듈에서 텍스처와 깊이 영상이 일대일 짝을 이루어 아틀라스 단위로 패킹된다. 생성된 아틀라스는 다중계층의 HEVC Main Profile 10으로 부호화 및 복호화되므로 기존의 부호화 시스템과 호환성이 보장된다.

2.1. 프루닝(Pruning)

그림 2는 TMIV 3.0에서 반영된 계층기반 프루너(Hierarchical Pruner)를 보인다[3]. 입력 시점 영상에서 기본 시점 영상(Basic view)는 Full view로 표기되어 어떤 화소도 제거되지 않는다. 아래의 Partial view는 부가 시점 영상(Additional view)로서 기본 시점 영상과 선순위의 부가 시점 영상들을 참조하여 중복된 영역을 제거하게 된다. 이때, 부가 시점 영상에서 제거되지 않는 화소들의 양상은 프루닝 순서에 따라 달라지며 이는 아틀라스 생성 시에 데이터 비트율과 렌더러에서의 합성 품질에 직접적으로 영향을 준다.

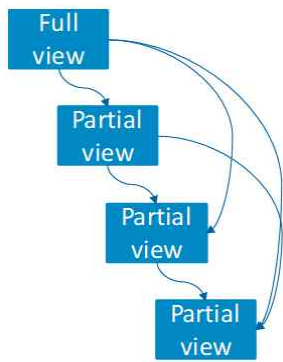


그림 2. TMIV 3.0 프루닝 방법

2.2. 최소 중첩 영역 기반 프루닝 순서 결정

본 논문에서 제안하는 프루닝 순서 결정 방법에 대한 가정은 다음과 같다. 기본 시점 영상과 부가 시점 영상 간의 중첩 영역(Overlapped area)이 적은 영역은 상대적으로 기본 시점 영상에서 보이지 않는 정보를 더 많이 가지고 있게 되므로 프루닝 후에 남은 잔차 화소들을 패킹할 때 패치의 크기가 크고, 패치의 수가 작게 되는 경향을 갖게 된다. 이에 따라, 기본 시점 영상과 중복되는 영역이 적은 순으로 우선순위를 부여하여 프루닝 순서를 결정한다[4].

그림 3은 최소 중첩 영역 기반 프루닝을 그림으로 보인다. 그림의 왼쪽 그룹에는 기본 시점 영상과 순위가 결정된 선순위의 부가 시점 영상이 있다. 부가 시점 영상에서 오른쪽 그룹의 각 후보 부가 시점 영상으로 3차원 워핑(warping)을 거쳐 중첩 영역을 계산하고 가장 적은 영역을 가진 부가 시점 영상이 왼쪽 그룹으로 이동하여 다음 순위를 갖게 된다. 이는 TMIV 3.0에서 채택된 계층기반 프루너와 최소 중첩 영역을 고려한다는 점에서 유사한 점이 있으나 프루닝 수행 방법에 있어서 차이가 있다.

TMIV 3.0의 계층기반 프루너는 목표 시점에서 중복 화소를 정할 때 실제로 래스터라이즈(rasterize)까지 수행하여 실제 렌더링 된 결과에서의 중첩 영역을 계산한하고, 또한 그림 4에서처럼 프루닝과 프루

닝 순서가 결합된 구조로서 프루닝을 수행하면서 프루닝 순서를 동시에 수행한다. 이는 시간복잡도 측면에서 이득이 있으나 프루닝 방법에 프루닝 순서가 종속된다는 단점이 있을 수 있다. 프루닝 순서는 다른 중요도에 의해 때로는 순서 변경이 요구되는 상황이 생길 수 있기 때문이다.

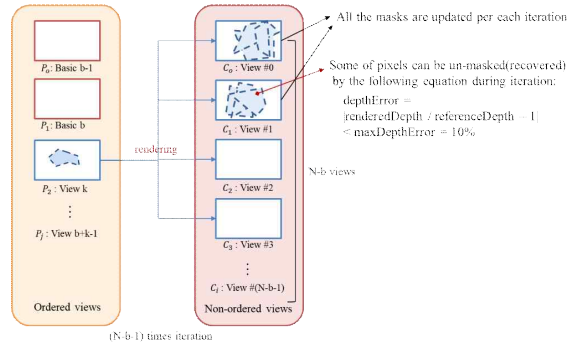


그림 3. 최소 중첩 영역 기반 프루닝 방법

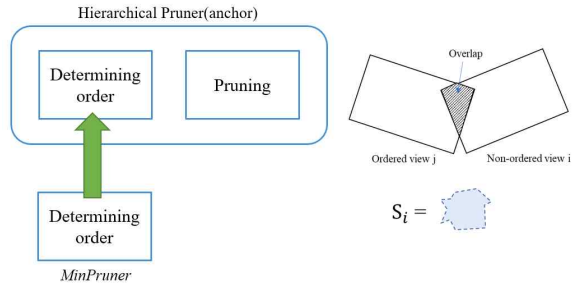


그림 4. 프루닝 순서 결정의 독립 구조

본 제안 방법은 기본적으로 프루닝 기법은 계층기반 프루너의 구조를 그대로 사용하고, 그림 4에서처럼 프루닝 순서 결정 과정을 중간에 두고 이를 최소 중첩 영역 기반의 프루닝 방법을 통해 순서를 넘겨주는 구조를 제안한다.

3. 실험 결과 및 분석

Group-based encoding	Sequence	Anchor (f)	High-BR BO rate Y-PSNR	Low-BR BO rate Y-PSNR	Max delta Y-PSNR	High-BR BO rate VMAF	Low-BR BO rate VMAF	High-BR BO rate MS-SSIM	Low-BR BO rate MS-SSIM	High-BR BO rate IV-PSNR	Low-BR BO rate IV-PSNR	Pixel rate ratio
Not applied	ClassroomVideo	AA97 (MV)	0.5%	0.4%	-4.26	1.8%	0.9%	1.0%	0.9%	0.4%	0.5%	0.00%
	TechnicalcolorMuseum	BA97 (MV)	-8.8%	-11.1%	11.71	-6.4%	-9.3%	-6.1%	-8.8%	-4.1%	-6.6%	0.00%
	TechnicalcolorHjck	CA97 (MV)	-10.6%	-7.1%	10.55	-5.7%	-4.8%	-11.4%	-6.5%	-14.7%	-9.6%	0.00%
	NokiaChess	NA97 (MV)	11.3%	5.1%	14.40	23.4%	13.7%	17.0%	9.5%	7.0%	1.5%	0.00%
Applied	M.V		-1.9%	-3.1%	14.40	3.3%	6.1%	0.1%	-1.2%	-2.8%	-3.8%	0.00%
	OrangeKitchen	JA97 (MV)	1.1%	1.4%	12.80	-0.3%	0.7%	-1.3%	-0.4%	-2.8%	-1.8%	0.00%
	TechnicalcolorPainter	TA97 (MV)	0.9%	0.8%	6.99	-0.2%	-0.1%	-0.3%	0.0%	1.7%	1.2%	0.00%
	IntellFrog	EA97 (MV)	0.7%	0.5%	8.90	0.7%	0.6%	1.3%	0.8%	1.4%	1.1%	0.00%
	PoznanFencing	LA97 (MV)	3.5%	2.9%	13.13	-6.7%	-4.7%	-1.3%	-0.4%	-0.7%	-0.6%	0.00%
	M.V		1.6%	1.4%	12.33	-1.6%	-0.9%	-0.4%	0.0%	-0.1%	0.0%	

그림 5. 정량적 결과 비교

	Pruning order
Anchor @ CG1-A	v0 v9 v12 v13 v14 v11 v10 v7 v8 v4 v3 v6 v1 v5 v2
MinPruning @ CG1-A	v0 v9 v12 v13 v11 v14 v8 v1 v2 v3 v4 v5 v6 v7 v10
Anchor @ CG1-N	v3 v7 v5 v1 v9 v0 v2 v4 v8 v6
MinPruning @ CG1-N	v3 v7 v9 v0 v5 v2 v4 v1 v6 v8

표 1. AA97과 NA97의 프루닝 순서 비교

본 제안 방법의 정량적 평가를 위해 MPEG-I Visual 그룹에서 채택 중인 평가 방법으로 총 8종의 실험영상에 대해 TMIV 3.0 앵커(Anchor)와 비교하였다. 그림 5는 8종의 실험영상을 그룹으로 나눈 결과를 보인다. 그룹으로 나뉘어 인코딩을 한 경우보다 비그룹 인코딩의 경우의 결

과가 대체적으로 개선이 되었음을 알 수 있다. 이는 그룹으로 나누어 인코딩을 하게 되면 그룹당 할당되는 시점의 수가 적어지기 때문에 프루닝 순서에 의한 영향이 적어지기 때문이다.

한 가지 주목할 것은 비그룹에서 AA97인 경우는 변화가 거의 없고, NA97인 경우만 유독 결과가 좋지 않다는 것이다. 분석을 위해 각각의 경우의 프루닝 순서를 표1로 정리하였다. AA97은 선순위에 해당하는 부가 시점 영상의 순서가 변함이 없었고, NA97은 9번과 0번 시점이 선순위로 지정되었다.



View 0 View 9
그림 6. 0번과 9번 시점 영상(NA97)

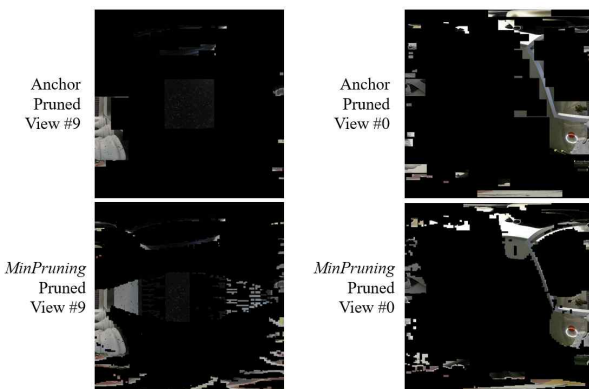


그림 7. 제안 기법의 0번과 9번 패치 비교(NA97)

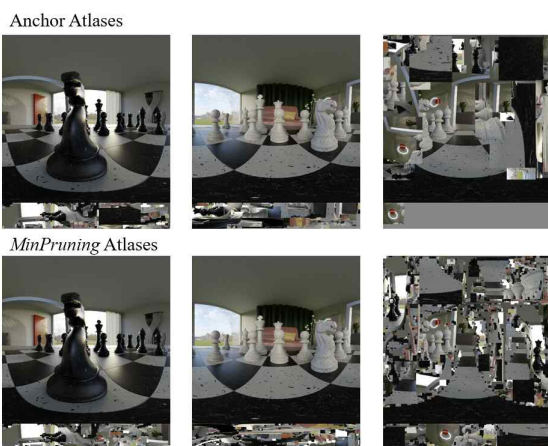


그림 8. 제안 기법의 아틀라스 비교(NA97)

NA97은 노키아(Nokia)에서 체스(Chess)를 CG로 생성한 실험영상으로 0번과 9번 시점은 다른 시점 카메라와 달리 그림 6에서처럼 360도 전방위에서 카메라 방향이 각각 위와 아래를 바라보고 있어서 최소 중첩 영역 기반 프루닝 기법을 적용했을 때 다른 시점과의 중첩되는 영역이 적어서 가장 높은 순위로 지정된다. 중첩 영역이 최소이지만 정반대 방향을 가리키는 시점이 선순위로 선택되어서 그림 7에서처럼 전체 장면

에서 비중이 적어야 하는 천장과 바닥 영역이 선순위로 지정되었고, 그 결과 프루닝 과정에서 장면에서 중요하지 않은 영역들이 선순위로 많이 남게 되어 그림 8에서처럼 아틀라스의 패치의 크기가 작아지고, 패치의 수가 많아지게 된 것으로 분석된다.

이 문제를 해결하기 위해 천장과 바닥과 같이 모든 카메라의 시점의 방향의 양상과 동떨어진 시점의 카메라를 검출하여 후순위로 자동 배정하도록 예외 처리 과정을 추가하여 개선할 수 있었고, 그림 9는 개선된 실험 결과를 보인다.

Sequence	Anchor (f)	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate MS-SSIM	Low-BR BD rate MS-SSIM	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate ratio
NokiaChess	NA97 (MIV)	11.3%	5.1%	14.40	23.4%	13.7%	17.0%	9.5%	7.0%	1.5%	0.00%
NokiaChess	NA97 (MIV)	0.7%	2.5%	14.96	9.0%	7.7%	3.4%	4.0%	5.9%	3.6%	0.00%

그림 9. 예외 처리를 통한 개선 결과(NA97)

4. 결론

본 논문에서는 MPEG MIV의 레퍼런스 소프트웨어인 TMIV 3.0에서 프루닝 순서를 결정하는 방법을 제안하였다. 기본 시점 영상과 부가 시점 영상 간의 중첩 영역(Overlapped area)이 적은 영역에 높은 우선 순위를 주는 방법으로 프루닝 순서를 결정하고 이 과정을 기존 방법인 계층 기반 프루닝에 중간 단계로서 프루닝 순서 결정 단계를 두어서 독립적인 구조로 바꾸었다. 결과로서 기존 방법보다 비그룹에서 정량적 평가수치에 향상이 있었다. 또한, 제한한 최소 중첩 영역 기법이 잘못 적용되는 예를 발견하여 원인을 확인하였고 프루닝 순서 결정 시에 예외처리를 통한 해결할 수 있었다.

감사의 글

본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(과제번호: 2018-0-00207, 이머시브 미디어 전문연구실).

참고문헌

- [1] 이광순, 정준영, 신홍창, 서정일, “이머시브 미디어 3DoF+ 비디오 부호화 표준 동향”, 전자통신동향분석, 제 34권, 제 6호, pp.156-163. 2019.
- [2] B. Salahieh, B. Kroon, J. Jung, and M. Domański, “Test Model 3 for Immersive Video,” ISO/IEC JTC 1/SC 29/WG 11 MPEG N18795, October 2019
- [3] B. Kroon and B. Sonneveldt, “Philips response to Immersive Video CE-2 (Pruner)”, ISO/IEC JTC1/SC29/WG11, M49662, October 2019.
- [4] 신홍창, 윤준영, 음호민, 이광순, 서정일, “이머시브 비디오 테스트 모델에서의 프루닝 순서 결정 기법 연구”, 제 32회 영상처리 및 이해에 관한 워크샵. 2020.