

딥러닝 기반 초해상화 기술을 이용한 MIV 성능 개선

*정태현 *이윤섭 **오관정 *오병태

*한국항공대학교 항공전자정보공학부, **한국전자통신연구원

*sdfds5566@kau.kr * yoonseob40@kau.kr **kjoh@etri.re.kr *byungoh@kau.kr

Improvement of MIV using Deep Learning based Super Resolution

*TaeHyun Jeong, *YoonSeob Lee, **Kwan-Jung Oh, *Byung Tae Oh

*Korea Aerospace University, **Electronics and Telecommunications Research Institute

요 약

본 논문에서는 TMIV 부호화 과정에서 개선된 압축성능을 위해 딥러닝을 이용한 초해상화 기술을 적용하는 방식을 제안한다. 제안 방식에서는 TMIV 인코더에서 아틀라스 생성한 후, 해당 아틀라스의 패킹된 뷰들을 downsampling하여 뷰들이 축소된 아틀라스를 생성하는 방식을 사용한다. 생성된 아틀라스는 기존의 방식 그대로 VVC를 이용하여 부복호화를 한다. 복호화된 아틀라스를 렌더링을 위해 뷰로 만드는 과정 중에 딥러닝을 이용한 초해상화 기술을 적용하여 줄어든 뷰들을 원래의 크기로 복원시킨다. 제안 기술을 통해 복원된 뷰의 화질을 유지시킨 채 많은 비트율을 감소시킬 수 있음이 확인된다.

1.서론

최근 메타버스에 대한 관심이 급증하고, VR과 같은 하드웨어가 발전하면서 단순한 2D 영상이 아닌 실감미디어 영상에 대한 수요가 증가하고 있다. 그 중 Immersive video는 사용자가 보는 시점에 따라 다양한 시점의 카메라에서 촬영된 영상들을 합성하여 보여줌으로써 보다 실감나는 영상을 시청할 수 있다. 그러나 이러한 Immersive video는 여러 시점에서 촬영된 영상의 데이터가 필요하기 때문에 시청하기위해 필요한 데이터가 일반적인 2d영상에 비해 매우 크다. 따라서 이러한 데이터에 대한 압축성능을 향상시키기 위해 MPEG에서 진행하고 있는 부호화 표준을

MIV (MPEG Immersive Video)라고 한다. MIV의 Test모델인 TMIV는 각 뷰를 적절하게 패킹하여 아틀라스를 생성하는 TMIV 인코딩 과정과, 생성된 아틀라스를 기존 2D 영상압축 기술로 부호화 하는 과정으로 압축을 진행한다. 아틀라스는 컬러 정보를 가진 texture 아틀라스와 깊이 정보를 가진 geometry 아틀라스가 있다. TMIV인코딩 방식은 2가지가 있다. MIV mode는 시점 간의 거리를 고려하여 선택된 basic view들과 시점 간의 중복성을 제거하기 위해 프루닝된 additional view를 아틀라스에 패킹하는 방식이고, MIV View mode는 basic view들만을 아틀라스에 패킹한다.

본 논문에서는 아틀라스에 패킹된 뷰 패치들을 downsampling 하여 압축된 데이터의 비트율을 낮추고, 복호화된 아틀라스에서 딥러닝 기반의 초해상화 기술을 적용하여 downsampling에 의해 손실된 정보를 최대한 복원하여 전통적인 모델 기반의 upsampling 방식보다 렌더링 성능을 향상시켰다.

2. 제안하는 기법

본 논문에서는 아틀라스에 basic view들만을 패킹하는 MIV view mode를 기반으로 실험을 진행했다. 시퀀스별로 texture와 geometry가 영향을 미치는 정도가 다르기 때문에 두 아틀라스 모두 sampling 기법을 적용하여 압축 성능을 높였다. 다만 딥러닝 기술을 적용하면 TMIV 연산량과 메모리소요가 크게 증가하기 때문에 정보량이 많은 texture 아틀라스에만 본 논문의 제안 기법



그림1. 제안 모델의 구조도

을 적용했고, geometry 아틀라스의 경우에는 기존 TMIV와 동일한 기술을 적용하였다.

2.1 Downsampling

부호화 성능을 향상시키기 위해 TMIV 인코더에서 아틀라스에 패킹된 각 뷰들의 texture를 전부 bicubic downsampling하였다. 그림 1과 같이 downsampling 이후 아틀라스는 비유효 영역이 증가하게 되는데, 이 영역은 비디오 코덱을 통해 압축하는 과정에서 완전한 중복성을 가지기 때문에 전체 아틀라스의 비트율은 감소하게 된다.

2.2 딥러닝 기반의 Upsampling

VVC의 부복호화 과정을 거치고 난 결과 아틀라스에 TMIV 디코더를 적용하기 전 우선 아틀라스 내부의 downsampling 된 저해상도 뷰들을 고해상도로 복원하기 위해 upsampling을 진행하였다. 사이즈를 고려하여 texture에만 딥러닝 기반의 upsampling이 적용된다. 초기 실험으로 downsampling과 동일하게 bicubic 방식을 이

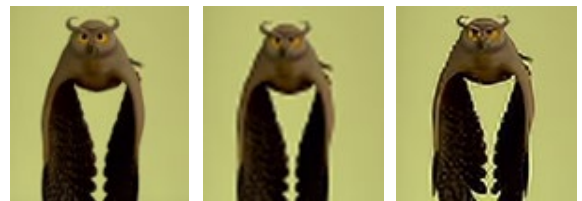
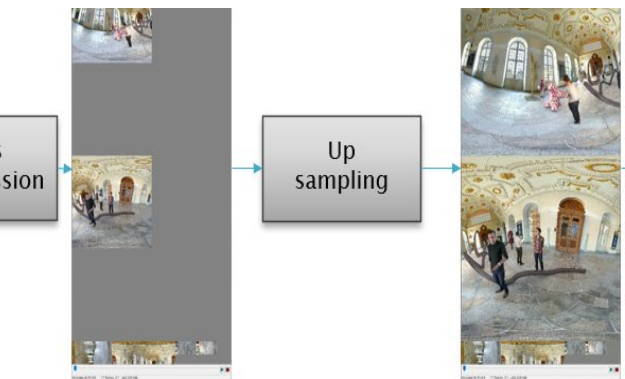


그림2. 좌측부터 Anchor, Bicubic, 딥러닝 기반 초해상화의 주관적 화질 비교



제안하는 방식의 성능을 평가하기 위하여, MIV mandatory sequence 중 B, D, E, J, O의 5개의 sequence에 대하여 실험을 진행하였고, 각 view 패치별로 bicubic과 EDSR로 얻어진 PSNR 비교 및 렌더링이 끝난 view의 WS-PSNR을 측정하여 성능을 비교하였다.

Seq	Bicubic	EDSR	Δ
B	34.33	34.59	0.26
D	45.73	45.79	0.06
E	38.26	37.28	-0.98

J	36.41	37.33	0.92
O	36.35	37.45	1.10
AVG	38.66	38.91	0.25

표1. 복원된 뷰의 화질 (PSNR) 비교

렌더링 성능을 평가하기 위한 WS-PSNR은 3D를 2D상으로 투영했을 때 생기는 왜곡을 가중치를 적용하여서 보정하는 평가지표이다. 원본 뷰와 TMIV 부복호화 과정을 거쳐 렌더링까지 완료된 뷰 사이의 WS-PSNR은 다음과 같이 계산된다.

$$w(i, j)_{ERP} = \frac{\cos(j+0.5-\frac{N}{2})\pi}{N} \quad (1)$$

Seq	QP1	QP2	QP3	QP4	QP5
B	0.06	0.03	0.01	0.00	0.00
D	0.02	0.01	0.01	-0.02	-0.02
E	-0.06	-0.05	-0.05	0.02	0.00
J	0.30	0.28	0.27	0.21	0.13
O	0.17	0.14	0.06	0.02	0.01

표2. 각 sequence의 QP별 bicubic대비 렌더링 성능 (WS-PSNR)

4. 결론

아틀라스에서 각 패치들을 downsampling을 하고 이후에 초해상화 기술로 복원해주는 방식을 적용하여, 비트율 저감에 있어서 상당한 이득을 얻을 수 있었으며, WS-PSNR은 자체는 anchor보다 낮아지긴 했지만, 비트율 대비 PSNR의 경우엔 대부분의 sequence에서 거의 anchor에 근접하거나 anchor보다 더 좋은 결과를 얻을 수 있었다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2018-0-00207, 이머시브전문 연구실)

참조문헌

- [1] Lim, Bee, et al. "Enhanced deep residual networks for single image super-resolution." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.
- [2] Salahieh, Basel, et al. "Test model 8 for MPEG immersive video." 133rd MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/WG4N0050. 2021.
- [3] H. T. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham and T. C. Thang, "An evaluation of quality metrics for 360 videos," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), 2017, pp. 7-11, doi: 10.1109/ICUFN.2017.7993736.
- [4] J. M. Boyce et al., "MPEG Immersive Video Coding Standard," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1521-1536, Sept. 2021, doi: 10.1109/JPROC.2021.3062590.