졸업논문
소개

6DoF 비디오 압축을 위한 점유맵 오류 보정 기법 및 신경망 기반 압축 성능분석

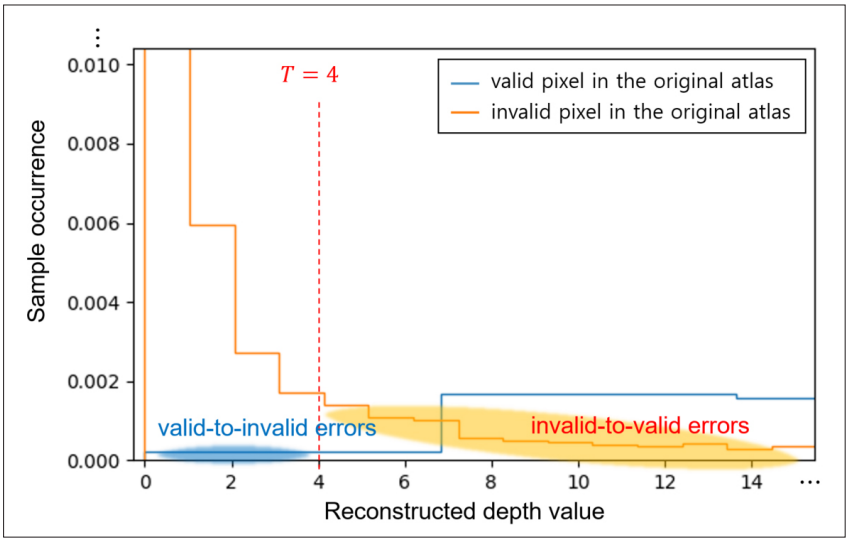
김동하 / 한국항공대학교 Media Communication Lab

3차원 공간에서의 전방위(omnidirectional) 시점과 함께 움직임 시차(motion parallax)를 제공하는 6자유도(6 Degree of Freedom: 6DoF)의 몰입형(immersive) 비디오는 메타버스 미디어의 핵심으로 더욱 주목받고 있다. MPEG(Moving Picture Experts Group)은 제한된 3D 공간의 여러 위치에서 획득한 다시점 비디오 및 깊이(Multiview plus Depth: MVD) 비디오로 구성된 방대한 데이터의 몰입형 비디오를 압축하기 위한 MIV(MPEG Immersive Video) 표준을 개발하였다. MIV는 MVD의 시점간 중복성을 제거한 아틀라스(atlas)를 기존의 2D 코덱을 이용하여 압축한다.

또한, 최근 MPEG 비디오 그룹은 6DoF 비디오 부호화를 위한 신경망(neural network) 기반의 새로운 접근방법에 주목하고 있다. 암시적 신경망 표현(Implicit Neural Representation: INR)을 이용하여 2D/3D 비디오를 신경망 모델로 표현하고 압축하는 암시적 시각 신경 표현(Implicit Neural Visual Representation: INVR)에 대한 표준화 가능성을 탐색하고 있다.

본 연구는 6DoF 몰입형 비디오의 보다 효율적인 부호화를 위하여 MIV의 아틀라스 생성 과정에서 시점간의 중복성 제거 후 텍스처(texture)와 깊이(depth)의 유효 화소 여부를 나타내는 점유맵(occupancy map)의 오류 보정 기법을 제안한다. 또한, 또 다른 6DoF 비디오 부호화 접근방법으로 6DoF 비디오를 대표적인 3D INR 모델인 NeRF(Neural Radiance Field) 모델로 학습한 다음 신경망 압축 표준인 NNC(Neural Network Compression) 코덱으로 학습된 모델을 압축하고 그 성능을 MIV의 참조 소프트웨어 코덱인 TMIV(Test Model for MIV)의 성능과 비교 분석한다.

MIV의 점유맵은 별도의 이진맵을 구성하여 압축하지 않고 깊이 정보에 내장하여 압축한다. 16비트의 깊이 정보는 아틀라스를 구성할 때 10비트의 깊이로 양자화되며, 이때 임계값 T를 사용하여 $2T$ ~ 1023의 값으로 양자화된다. 이때, 0 ~ $2T$ 구간에 점유 정보가 내장되는데 T 보다 작은 값은 비유효 화소, T 보다 큰 값은 유효 화소임을 나타낸다. 하지만, 깊이 아틀라스의 압축 오류로 인하여 〈그림 1〉과 같이 내장된 점유 정보에 오류가 발생할 수 있다.

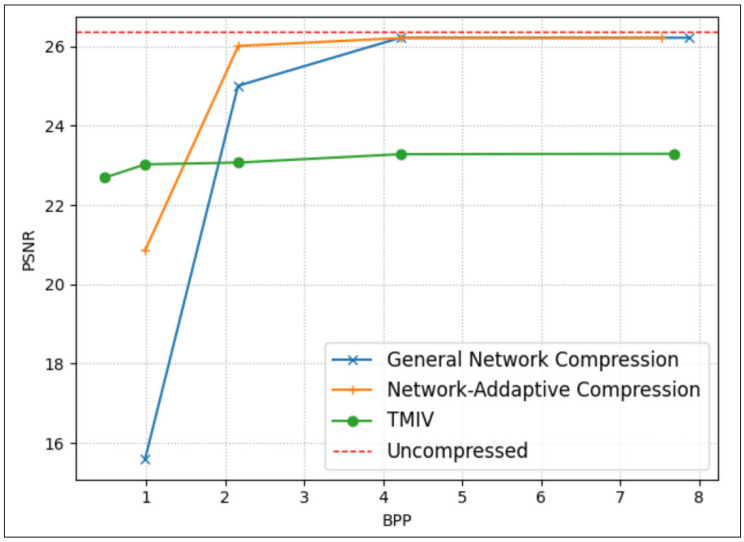


<그림 1> 복호화된 깊이 아틀라스의 히스토그램(Fan, T = 4, QP 5) 및 점유맵 오류

제안하는 점유맵 오류 보정 기법은 기존에 TMIV에 설정된 T 값을 조정하여 점유 오류 보정과 함께 렌더링 화질을 개선하여 부호화 효율을 개선한다. 즉, T 값을 기존보다 작게 설정하면 깊이 정보를 표현하는 동적범위가 확장되어 보다 정확한 깊이 정보를 통해 렌더링 화질을 개선할 수 있다. 반면, 낮춰진 임계값으로 인해 압축과정에서 비유효 화소가 유효 화소로 변경되는 점유 오류가 증가하게 되고 이로 인하여 렌더링 뷰에 많은 시각적 아티팩트(artifact)가 발생한다. 따라서, 제안방법은 MIV의 부호화기와 복호화기에서 점유 정보의 유효 여부를 판정하는 임계값을 비대칭적으로 사용하여 점유 정보 오류 보정과 렌더링 화질을 동시에 개선한다. 실험결과 부호화기에서 T=4, 복호화기에서 T=6으로 설정했을 때 렌더링 화질을 유지하며 많은 시각적 아티팩트를 감소시켰다. 제안방법을 깊이의 정확도가 높은 CG 콘텐츠에만 적용한 경우 평균 2.2%의 BD-rate 이득 성능을 보였으며 이러한 성능 검증을 바탕으로 MIV Edition 2 WD(Working Draft)와 TMIV 15.0에 채택되었다.

신경망 기반의 6DoF 비디오 압축 연구에서는 다시점 비디오와 카메라 정보를 사용하여 NeRF 모델을 학습하고 학습된 모델을 NNC로 압축한다. NeRF는 렌더링되는 화소에서 나아간 레이(ray)의 구간별 샘플들의 밀도와 색상값을 MLP(Multi-Layer Perceptron)를 사용하여 예측하고 이 샘플들의 값을 합산하여 렌더링된 화소의 색상값을 결정함으로써 MIV와 같이 임의의 시점에서의 비디오를 렌더링하다. 이때, NeRF는 MLP로 구성된 성긴(coarse) 네트워크와 미세(fine) 네트워크를 사용하여 샘플의 밀도와 색상값을 예측한다. 본 연구에서는 NeRF로 다시점의 비디오를 학습하고 MLP로 구성된 두 네트워크를 NNC 압축함으로써 신경망 기반의 몰입형 비디오의 부호화를 수행하고 그 성능을 TMIV의 성능과 비교 분석했다.

NNC를 활용하여 두 네트워크를 압축할 때, 전처리 과정인 파라미터 감축은 제외하고 양자화와 엔트로피 부호화를 포함한다. 이때, NeRF의 두 네트워크의 중요도에 따라 차별적으로 압축 비트량을 할당하는 NeRF 네트워크-적응적 비트할당(network-adaptive bit allocation) 압축 기법을 제안한다. 성긴 네트워크보다 미세 네트워크가 레이 상의 밀도와 색상값 추론에 더 중요한 역할을 한다. 따



<그림 2> NeRF 네트워크-적응적 비트할당 부호화 성능(Mirror)

라서, 미세 네트워크에 상대적으로 더 작은 양자화 파라미터(QP) 값을 적용하여 많은 비트량을 할당하였을 때, 동일 압축 비트율에서 개선된 렌더링 성능을 유지한다.

〈그림 2〉는 NNC를 사용하여 동일한 QP로 두 네트워크를 압축했을 때, 제안방법의 압축율(BPP) 대비 렌더링 화질(PSNR)을 TMIV와 비교한 것이다. 입력 다시점의 비디오 중 두 개의 뷰를 제외하고 NeRF 학습 및 압축을 수행하였고, 렌더링 성능은 제외된 두 개의 뷰에 대한 결과이다. 공정한 성능비교를 위해 TMIV도 동일하게 두 개의 뷰를 제외하고 부호화를 수행하였고 제외된 두 개의 뷰에 대한 렌더링 화질을 확인하였다. 제안기법이 동일한 압축율에서 TMIV보다 높은 화질로 학습에 사용되지 않은 새로운 뷰를 렌더링할 수 있음을 확인하였다. 제안기술은 MPEG INVR 표준화에 기고되었으며 표준 탐색기술로 검토되었다.

김 동 하

- 2021년 8월 : 한국항공대학교 항공전자정보공학부 학사
- 2023년 8월 : 한국항공대학교 항공전자정보공학과 석사
- 2023년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학과 박사과정
- 주관심분야 : 딥러닝, 몰입형 비디오 부호화(MPEG MIV, INVR)