Check for updates
일반논문 (Regular Paper)
방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)
https://doi.org/10.5909/JBE.2026.31.5.1052
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

자유시점 비디오를 위한 가우시안 스플래팅 압축에서의 잠재 표현 희소화 기법

김 태 유[a)], 박 장 수[a)], 심 동 규[a)‡]

Latent Sparsification for Gaussian Splatting Compression in Free-Viewpoint Video

Taeyu Kim[a)], Jangsoo Park[a)], and Donggyu Sim[a)‡]

요 약

본 논문에서는 동적 가우시안 스플래팅 기반 온라인 자유시점 비디오 (FVV)의 압축 효율 향상을 위한 잠재 표현 희소화를 적용하고, 그 효과를 실험적으로 분석한다. 제안하는 방법은 양자화 이전 잔차 잠재 표현에 L1 정규화를 적용하여 잠재 표현을 0 주변으로 집중시키고, 양자화 이후 0 심볼의 출현 비율을 높여 엔트로피 부호화 효율을 향상시킨다. 또한 기존 압축 구조를 변경하지 않고 손실 함수에 정규화 항만 추가하므로 구현이 간단하며 기존 압축 파이프라인과 호환성을 유지한다. N3DV와 MPEG GSC 데이터셋 실험에서 제안하는 방법은 기존 QUEEN 대비 PSNR 손실 0.3 dB 이내에서 평균 55.4%, 최대 68.4%의 잔차 저장 용량 절감을, 표준편차 기반 정규화 대비 평균 7.4%의 BD-rate 개선을 달성하였다. 정규화 강도에 따른 잠재 표현의 분포 분석 결과, L1 정규화가 잠재 표현을 압축에 유리한 형태로 희소화함을 확인하였다.

Abstract

This paper applies latent sparsification to improve the compression efficiency of dynamic Gaussian splatting-based online free-viewpoint video (FVV) and experimentally analyzes its effect. The proposed method employs L1 regularization on pre-quantized residual latents, driving the latent representations toward zero and thereby increasing the proportion of zero-valued symbols after quantization. As a result, the entropy of quantized latents is reduced, leading to improved entropy coding efficiency. The proposed method requires only a single regularization term in the training objective without modifying the existing compression architecture, making it readily applicable to existing compression pipelines. Experimental results on the N3DV and MPEG GSC datasets show that the proposed method achieves an average residual storage reduction of 55.4% and a maximum reduction of 68.4% over QUEEN with less than 0.3 dB PSNR loss, as well as an average 7.4% BD-rate reduction over standard deviation-based regularization. Analysis of the latent distribution across regularization strengths confirms that L1 regularization sparsifies latent representations into a form more favorable for compression.

Keyword : Free-Viewpoint Video, Dynamic Gaussian Splatting, Video Compression, Latent Sparsification, Entropy Coding

Copyright Ⓒ 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.
"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

Ⅰ. 서 론

자유시점 비디오 (Free-Viewpoint Video, FVV)는 사용자가 원하는 시점과 시간에서 장면을 자유롭게 관찰할 수 있는 차세대 몰입형 미디어 기술로, 실감형 방송, 가상·증강현실 (AR·VR), 디지털 트윈, 원격 협업 등 다양한 응용 분야에서 활용되고 있다. 기존 영상이 촬영된 카메라 시점만을 제공하는 것과 달리, FVV는 임의의 시점에서 새로운 영상을 생성할 수 있어 높은 수준의 몰입감을 제공한다. 그러나 이러한 서비스를 위해서는 장면을 구성하는 대규모 3차원 정보를 저장하고 전송해야 하므로 높은 저장 공간과 전송 대역폭이 요구된다. 따라서 실제 FVV 서비스를 위해서는 고품질 콘텐츠를 유지하면서도 저장 용량과 전송 비용을 효과적으로 감소시키는 압축 기술이 필수적이다[1,2].

최근에는 가우시안 스플래팅 (Gaussian Splatting)을 이용한 장면 표현 기술의 발전으로 FVV 콘텐츠 생성 기술이 크게 발전하였다. 초기 Neural Radiance Field (NeRF)[3]는 다중 시점 영상으로부터 고품질 Novel View Synthesis를 가능하게 하였으나 높은 계산 복잡도로 인해 실시간 응용에는 한계가 있었다. 이후 제안된 3D 가우시안 스플래팅 (3DGS)[4]은 장면을 명시적인 가우시안 집합으로 표현하여 높은 렌더링 품질과 실시간 처리 속도를 동시에 달성하였으며, 최근에는 동적 가우시안 스플래팅 (Dynamic Gaussian Splatting) 기술[5]이 제안되면서 시간에 따라 변화하는 장면까지 효과적으로 표현할 수 있게 되었다. 그러나 동적 가우시안 스플래팅은 수많은 가우시안의 기하 정보와 속성 정보를 저장해야 하므로 데이터 크기가 매우 크며, 이를 효율적으로 저장하고 전송하기 위한 압축 기술의 중요성이 더욱 커지고 있다.

이러한 문제를 해결하기 위하여 최근에는 가우시안 스플래팅 압축 기술이 활발히 연구되고 있다[6-8]. 정적 장면에서는 가우시안 프루닝 (Pruning), 양자화 (Quantization) 및 엔트로피 부호화 (Entropy Coding)를 이용하여 저장 용량을 감소시키는 다양한 방법들이 제안되었으며, 동적 장면에서는 프레임 간 중복성을 활용하는 잔차 부호화 (Residual Coding) 기반 압축 기술들이 연구되고 있다. 특히 QUEEN (Quantized Efficient Encoding of Dynamic Gaussians)[9]은 프레임 간 가우시안 속성 잔차 (Attribute Residual)를 저차원 잠재 표현 (Latent Representation)으로 변환한 후 양자화와 엔트로피 부호화를 수행하는 구조를 통해 우수한 압축 성능을 달성하였으며, 동적 가우시안 스플래팅 기반 온라인 FVV 압축의 대표적인 방법으로 활용되고 있다. 하지만 기존 가우시안 스플래팅 압축 기술은 대부분 잠재 표현을 얼마나 효율적으로 생성하고 표현할 것인가에 초점을 맞추고 있으며, 예측 네트워크 (Prediction Network), 잠재 변환 (Latent Transform), 양자화 및 엔트로피 모델 등을 개선하여 압축 성능을 향상시켜 왔다.

일반적으로 엔트로피 부호화의 평균 비트 수는 양자화된 잠재 심볼 (Latent Symbol)의 확률 분포에 의해 결정되며, 심볼 분포 (Symbol Distribution)의 엔트로피가 낮을수록 필요한 평균 비트 수가 감소한다. 즉 심볼의 확률 분포가 일부 심볼에 집중될수록 엔트로피가 감소하여 더 높은 압축 효율을 얻을 수 있다. 한편, 양자화된 잠재 심볼은 양자화 이전의 실수 잠재 표현으로부터 생성되므로, 실수 잠재 표현의 분포는 양자화 이후 심볼 분포의 형성에 직접적인 영향을 미친다. 따라서 학습 과정에서 실수 잠재 표현의 분포를 적절히 제어한다면 기존 압축 구조를 변경하지 않고도 엔트로피를 효과적으로 감소시켜 압축 효율을 향상시킬 수 있다. 그러나 기존 연구들은 대부분 잠재 표현의 표현 능력 향상이나 양자화 이후의 과정에 집중하고 있으며, 별도의 엔트로피 모델을 추가하지 않고 실수 잠재 분포를 직접 제어하여 엔트로피 부호화 효율을 향상시키는 접근은 충분히 연구되지 않았다.

이를 해결하기 위하여 본 논문에서는 양자화 이전 실수 잠재 분포를 제어하기 위해 L1 기반 잠재 표현의 희소화 (Latent Sparsification) 방법을 적용하고, 이에 따른 압축 효율과 잠재 표현의 분포 변화를 실험적으로 분석한다. 제안

a) 광운대학교 컴퓨터공학과(Department of Computer Engineering, Kwangwoon University)
‡ Corresponding Author : 심동규(Donggyu Sim)
E-mail: dgsim@kw.ac.kr
Tel: +82-2-940-6470
ORCID: https://orcid.org/0000-0002-2794-9932
※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원(IITP-2026-RS-2022-00156225) 및 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-학·석사연계 ICT 핵심인재양성의 지원(IITP-2026-RS-2022-00156215)을 받아 수행된 연구임.
· Manuscript received July 20, 2026; Revised August 24, 2026; Accepted August 24, 2026.

하는 방법은 학습 과정에서 실수 잠재 표현의 절댓값을 감소시켜 양자화 이후 0으로 매핑되는 값의 비율을 증가시킴으로써, 0 심볼의 출현 빈도를 높이고 엔트로피를 감소시켜 평균 비트 수를 효과적으로 줄인다. 또한 제안하는 방법은 기존 QUEEN의 인코더, 디코더, 양자화 및 엔트로피 부호화 구조를 변경하지 않고 손실 함수에 정규화 항 (Regularization Term)만 추가하여 적용할 수 있으므로 구현이 간단하고 기존 압축 파이프라인과 높은 호환성을 유지한다. N3DV 및 MPEG GSC 데이터셋을 이용한 실험 결과, 제안하는 방법은 PSNR 손실 0.3 dB 이내에서 기존 QUEEN 대비 평균 55.4%, 최대 68.4%의 잔차 저장 용량을 절감하였으며, 분포 분석을 통해 정규화 강도에 따라 잠재 표현이 압축에 유리한 형태로 희소화됨을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 가우시안 스플래팅 압축 기술과 관련 연구를 소개한다. 3장에서는 제안하는 L1 기반 잠재 표현 희소화 방법을 설명한다. 4장에서는 실험 결과와 성능을 분석하며, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 관련 연구

3D 가우시안 스플래팅[4]은 기존 NeRF[3] 기반 장면 표현 방식보다 높은 렌더링 품질과 실시간 처리 속도를 제공하면서 자유시점 비디오 생성 분야의 대표적인 장면 표현 기법으로 자리 잡았다. 그러나 장면을 구성하는 수많은 가우시안의 기하 정보와 속성 정보를 저장해야 하므로 높은 저장 비용이 요구되며, 이를 해결하기 위한 다양한 압축 기술이 제안되어 왔다. 최근에는 정적 장면을 대상으로 하는 가우시안 압축뿐 아니라 동적 장면을 위한 온라인 자유시점 비디오 압축과 학습 기반 엔트로피 모델링까지 연구가 확장되고 있다. 본 장에서는 가우시안 스플래팅 압축 기술을 정적 압축, 동적 압축, 그리고 희소 표현과 엔트로피 모델링으로 구분하여 살펴본다.

1. 정적 3DGS 압축

정적 가우시안 스플래팅 압축은 크게 가우시안의 개수를 감소시키는 방법, 가우시안 속성을 효율적으로 표현하는 방법, 그리고 율-왜곡 최적화를 통해 압축률을 명시적으로 제어하는 방법으로 구분할 수 있다. 첫 번째 접근은 장면을 표현하는 가우시안의 개수를 줄여 저장 공간을 감소시키는 방법이다. 대표적으로 Compact 3DGS[6]는 학습 가능한 마스크를 이용하여 가우시안의 중요도를 학습하고, 중요하지 않은 가우시안을 제거하는 학습 기반 프루닝 기법을 제안하였다. 또한 잔차 벡터 양자화를 적용하여 일부 속성을 효율적으로 압축하였다. 두 번째 접근은 가우시안 속성을 저차원 잠재 표현으로 변환하여 저장 효율을 높이는 방법이다. EAGLES[7]는 학습 가능한 양자화를 이용하여 가우시안 속성을 잠재 표현으로 변환하고, 학습 과정에서 압축 성능과 재구성 성능을 동시에 향상시키는 구조를 제안하였다. 세 번째 접근은 엔트로피 모델을 이용하여 비트량을 추정하고 이를 손실 함수에 반영하는 방법이다. CAT-3DGS[10]는 문맥 적응적 엔트로피 모델을 이용하여 가우시안 속성 간의 상관관계를 활용하고, 율-왜곡 최적화를 통해 압축 성능을 향상시켰다. 이러한 연구들은 가우시안의 개수나 속성 표현을 효율화하거나 율-왜곡 최적화를 도입함으로써 정적 장면의 저장 효율을 향상시켰다. 그러나 이들은 시간에 따라 변화하는 장면을 고려하지 않으므로, 동적 자유시점 비디오 압축에 직접 적용하기에는 한계가 있다.

2. 동적 3DGS 압축

동적 가우시안 스플래팅 압축은 전체 시퀀스를 대상으로 학습하는 오프라인 방식과 프레임을 순차적으로 처리하는 온라인 방식으로 구분할 수 있다. 오프라인 방식의 대표적인 연구인 4D 가우시안 스플래팅[5]은 기존 가우시안 집합을 유지하면서 시간에 따른 가우시안의 위치와 속성 변화를 신경망으로 예측하는 방법을 제안하였다. 그러나 전체 시퀀스를 동시에 학습해야 하므로 온라인 스트리밍 환경에는 적용하기 어렵다. 온라인 방식에는 이전 프레임 정보를 이용하여 현재 프레임을 점진적으로 표현하는 연구가 수행되었다. 3DGStream[11]은 Neural Transformation Cache를 이용하여 위치와 회전 변화를 예측하는 구조를 제안하였으나, 일부 속성만을 대상으로 하므로 표현 가능한 변화에 제약이 존재한다. QUEEN[9]은 이러한 한계를 개선하기 위하

여 위치를 제외한 모든 가우시안 속성을 이전 프레임과의 잔차 형태로 표현하고, 이를 잠재 표현으로 변환하여 양자화 및 엔트로피 부호화를 수행하는 점진적 압축 구조를 제안하였다. 또한 위치 정보는 학습 가능한 게이트를 이용하여 선택적으로 전송함으로써 압축 효율을 향상시켰다. 이는 변환, 양자화, 엔트로피 부호화로 이어지는 전통적인 변환 부호화 파이프라인과 동일한 구성에 해당한다.

이후 최근 연구들은 전송이 필요한 요소 자체를 줄이는 방향으로 발전하였다. ComGS[12]는 장면 모션의 국소성을 이용하여 소수의 키포인트만으로 전체 가우시안의 움직임을 표현하고 키포인트 속성만 전송함으로써 저장 용량을 크게 감소시켰다. ReCon-GS[13]는 계층적 앵커 가우시안을 밀도 적응적으로 배치하여 변형을 표현하고, 저장량을 고려한 최적화를 통해 재구성 속도와 압축 효율을 동시에 향상시켰다. 다만 이러한 방법들은 표현 구조 자체를 새롭게 설계해야 하므로 기존 압축 파이프라인과의 호환성에는 제약이 존재한다. 반면 QUEEN은 앞서 설명한 변환 부호화 구조를 따르므로 기존 파이프라인과의 호환성이 높으며, 온라인 자유시점 비디오 압축의 대표적인 기준 구조(Baseline)로 활용되고 있어 본 논문에서도 이를 기준 구조로 사용한다.

3. 희소 표현과 엔트로피 모델링

신호를 희소한 형태로 표현하여 압축 효율을 높이는 접근은 변환 부호화 기반 영상 압축에서 오랫동안 활용되어 온 원리이다. H.264/AVC[14]와 H.265/HEVC[15]를 비롯한 변환 부호화 기반 비디오 코덱은 변환을 통해 신호의 에너지를 소수의 변환 계수에 집중시키고, 양자화 이후 대부분의 변환 계수가 0이 되는 특성을 이용하여 0이 아닌 계수의 위치와 값만을 선택적으로 부호화한다[16]. 또한 통계적 학습 분야에서는 L1 정규화가 계수의 절댓값을 직접 축소시켜 희소한 해를 유도한다는 점이 알려져 있다[17]. 즉 희소성은 엔트로피 부호화 효율과 직접적으로 연결되는 성질이며, 본 논문은 이러한 원리를 동적 가우시안 스플래팅의 잠재 표현에 적용한다.

신경망 기반 압축에서는 잠재 표현의 확률 분포를 모델링하여 압축 효율을 향상시키기 위한 다양한 엔트로피 모델링 기법이 연구되어 왔다. 대표적으로 Ballé et al.[18]은 잠재 표현의 확률 분포를 학습하고 비트 수와 재구성 왜곡을 동시에 최적화하는 학습 기반 압축 구조를 제안하였다. 이후 Hyperprior[19]와 Context Model[20] 등이 제안되면서 잠재 표현의 확률 분포를 보다 정밀하게 추정하여 압축 성능을 향상시키는 방향으로 발전하였다. 이러한 엔트로피 모델링은 최근 동적 가우시안 스플래팅 압축에도 적용되고 있다. iFVC[21]는 이진 양자화된 잠재 표현의 경험적 발생 빈도로 비트량을 추정하여 율-왜곡 최적화를 수행하였고, 4DGC[22]는 모션 그리드와 보정 가우시안에 대해 암시적 엔트로피 모델을 학습하여 율-왜곡을 공동 최적화하였다. 그러나 이러한 방법들은 압축률을 명시적으로 추정하기 위해 별도의 엔트로피 모델이나 비트량 추정 과정을 학습 구조에 포함해야 한다. 따라서 온라인 자유시점 비디오와 같이 빠른 압축 및 복원이 요구되는 환경에서는 학습 비용과 저장 비용이 증가할 수 있다. 반면 본 논문에서는 별도의 엔트로피 모델을 추가하지 않고 기존 QUEEN 구조를 그대로 유지하면서 양자화 이전 잠재 표현의 분포를 직접 제어하여 엔트로피 부호화 효율을 향상시키는 방법을 제안한다.

Ⅲ. 제안하는 잠재 표현 희소화 방법

자유시점 비디오 압축에서 엔트로피 부호화 효율은 양자화된 잠재 심볼의 분포에 의해 결정된다. 그러나 기존 QUEEN은 재구성 품질과 위치 게이트만을 고려하여 학습되므로, 양자화 이전 잠재 표현의 분포는 압축 효율을 고려하지 않은 채 최적화된다. 본 장에서는 먼저 QUEEN의 전체 압축 구조를 설명한 후, 잠재 분포와 압축 효율의 관계를 통하여 기존 방법의 한계를 분석한다. 마지막으로 이러한 문제를 해결하기 위한 L1 기반 잠재 표현 희소화 방법을 제안한다.

1. QUEEN 기반 압축 구조

QUEEN[9]은 동적 가우시안 스플래팅 기반 자유시점 비디오를 위한 온라인 압축 방법으로, 프레임 간 시간적 중복성을 이용하여 압축 효율을 향상시킨다. 이는 각 프레

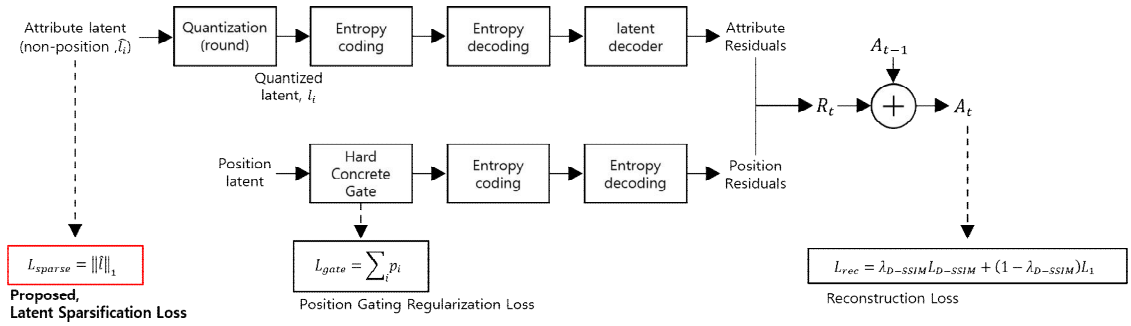

그림 1. 제안하는 L1 기반 잠재 표현 희소화 정규화 (빨강)를 적용한 QUEEN 압축 파이프라인 개요
Fig. 1. Overview of the QUEEN compression pipeline with the proposed L1-based latent sparsification loss (in red)

임을 독립적으로 표현하는 대신, 현재 프레임의 가우시안 속성을 이전 프레임에 대한 속성 잔차 (Attribute Residual) 형태로 표현하여 저장해야 하는 정보를 최소화한다. 또한 위치 정보와 위치를 제외한 속성 정보를 서로 다른 방식으로 압축함으로써 재구성 품질과 압축 효율을 동시에 고려한다.

[그림 1]은 QUEEN의 전체 압축 파이프라인과 본 논문에서 제안하는 잠재 표현 희소화 (L_{sparse})의 적용 위치를 함께 나타낸다. 위치를 제외한 속성은 잠재 표현으로 변환된 후 양자화와 엔트로피 부호화를 거쳐 전송되며, 위치 정보는 Hard Concrete Gate[23]를 이용하여 선택적으로 전송된다. 본 논문에서 제안하는 희소화 정규화는 위치를 제외한 속성의 실수 잠재 표현에만 적용되며, 디코더와 엔트로피 부호화 구조는 기존 QUEEN과 동일하게 유지된다.

QUEEN은 시점 t의 가우시안 속성(A_t)을 이전 프레임의 속성(A_{t-1})과 속성 잔차(R_t)의 합으로 표현한다.

$$A_t = A_{t-1} + R_t \tag{1}$$

속성 잔차는 위치 잔차와 위치를 제외한 속성 잔차로 구분되며, 두 정보는 서로 다른 특성을 가지므로 각각 다른 압축 방식을 적용한다. 위치 잔차는 양자화 오차에 매우 민감하며 작은 오차만으로도 렌더링 품질이 크게 저하될 수 있다. 따라서 QUEEN은 Hard Concrete Gate를 이용하여 전송이 필요한 가우시안만 선택하고, 선택된 위치 잔차를 실수 형태로 직접 전송한다. 반면 색상, 스케일, 회전 및 불투명도와 같은 속성 잔차는 양자화에 비교적 강인하므로, 학습된 실수 잠재 표현 $\hat{l}_i$로 변환한 뒤 정수 잠재 표현 l_i로 양자화하여 저장한다.

양자화 과정은 반올림 연산을 포함하므로 직접적인 역전파가 불가능하다. 이를 해결하기 위해 QUEEN은 Straight-Through Estimator (STE)를 사용하여 순전파에는 반올림 연산을 수행하고, 역전파에는 해당 연산을 항등 함수 (Identity Function)로 근사한다. 양자화된 정수 잠재 표현은 속성별 선형 디코더 D를 통해 실제 속성 잔차 r_i로 복원된다.

$$l_i = \mathrm{STE}(\hat{l}_i), \qquad r_i = D \cdot \mathrm{float}(l_i) \tag{2}$$

학습이 종료되면 정수 잠재 표현 (l_i)의 실제 심볼 분포를 이용하여 경험적 확률 분포를 구성하고, 이를 기반으로 산술 부호화를 수행하여 비트스트림을 생성한다. 복원 과정에서는 엔트로피 디코딩을 통해 정수 잠재 표현을 복원한 뒤, 속성별 디코더를 이용하여 속성 잔차를 복원하고 식 (1)에 따라 현재 프레임의 가우시안 속성을 재구성한다. 이때 별도의 학습된 엔트로피 모델을 사용하지 않고 정수 잠재 표현의 실제 심볼 분포를 직접 이용하므로, 압축 효율은 정수 잠재 표현 l_i의 심볼 분포에 직접적으로 영향을 받는다.

QUEEN의 학습은 재구성 품질과 전송량을 동시에 고려하는 목적 함수로 수행된다. 재구성 손실 (L_{rec})은 렌더링 영상과 원본 영상 사이의 픽셀 단위 L1 손실과 D-SSIM 손실의 가중합으로 정의되며, 전체 손실 함수 (L_{QUEEN})는 재구성 손실과 위치 게이트 손실 (L_{gate})을 함께 최적화하는 형태로 구성된다.

$$L_{QUEEN} = L_{rec} + \lambda_{gate} L_{gate} \tag{3}$$

2. 잠재 분포와 압축 효율 분석

양자화 기반 압축에서 엔트로피 부호화의 평균 비트 수는 양자화된 정수 잠재 표현의 심볼 분포에 의해 결정된다. 일반적으로 심볼의 확률 분포가 일부 심볼에 집중될수록 엔트로피는 감소하며, 이에 따라 평균 비트 수도 함께 감소한다. 특히 동적 자유시점 비디오에서는 프레임 간 변화가 상대적으로 작으므로 속성 잔차의 대부분은 0에 가까운 값을 갖는다. 따라서 정수 잠재 표현이 0 주변에 집중될수록 0 심볼의 출현 비율 (zero-ratio)이 증가하고, 엔트로피 부호화 효율 또한 향상된다. 그러나 식 (3)의 QUEEN 손실 함수는 재구성 손실과 위치 게이트 손실만을 최적화하며, 위치를 제외한 속성의 실수 잠재 표현의 분포를 직접 제어하는 항은 포함하지 않는다. 그 결과 실수 잠재 표현은 재구성 품질만을 기준으로 학습되며 압축 효율을 고려한 분포로 유도되지 않는다. 이후 수행되는 양자화와 엔트로피 부호화는 이미 학습된 잠재 표현을 이용하여 수행되므로, 학습 과정에서 압축에 유리한 심볼 분포를 형성할 기회가 존재하지 않는다.

이러한 현상은 실제 학습된 정수 잠재 표현의 심볼 분포에서도 확인할 수 있다. [표 1]은 N3DV의 coffee_martini, sear_steak과 MPEG GSC의 bartender, cinema 시퀀스에 대해 QUEEN 기준 구조를 학습한 후 잔차 프레임에서 생성된 정수 잠재 표현의 심볼 분포를 분석한 결과이다. 각 속성에 대해 심볼의 종류 수 (# Symbols), 0 심볼의 출현 비율 (zero-ratio), 그리고 심볼당 평균 비트 수 (bit/symbol)를 계산하였다. [표 1]에서 색상과 스케일 속성은 0 심볼의 출현 비율이 0.90 이상으로 높게 나타나며 심볼당 평균 비트 수도 0.2~0.9 bit 수준으로 낮다. 반면 불투명도 속성은 모든 시퀀스에서 0 심볼의 출현 비율이 약 0.5 수준에 머물고 심볼의 종류 수도 가장 많으며, 심볼당 평균 비트 수 역시 약 2 bit로 가장 크게 나타난다. 예를 들어 cinema 시퀀스에서 스케일 속성은 0 심볼의 출현 비율이 0.952로 심볼당 평균 0.321 bit만을 사용하는 반면, 불투명도 속성은 0 심볼의 출현 비율이 0.551에 불과하여 평균 2.008 bit가 필요하다. 이는 동일한 압축 구조에서도 잠재 표현의 분포가 달라짐에 따라 심볼 분포와 엔트로피 부호화 효율이 크게 달라질 수 있음을 보여준다.

표 1. QUEEN 기준 구조의 속성별 정수 잠재 표현 심볼 분포
(속성별 잠재 표현의 차원은 Color(DC)·Scale 8, Color(AC) 4, Rotation 6, Opacity 3이다.)
Table 1. Symbol distribution of integer latents per attribute in the QUEEN baseline
(The per-attribute latent dimensions are 8 for color(DC) and scale, 4 for color(AC), 6 for rotation, and 3 for opacity.)

Dataset	Attribute	# Symbols	zero-ratio	bit/symbol
coffee_martini (N3DV)	Color (DC)	5.0	0.961	0.278
	Color (AC)	9.7	0.850	0.797
	Scale	5.6	0.981	0.151
	Rotation	7.3	0.828	0.840
	Opacity	15.6	0.610	1.721
sear_steak (N3DV)	Color (DC)	5.3	0.959	0.286
	Color (AC)	11.9	0.847	0.814
	Scale	4.5	0.980	0.158
	Rotation	7.4	0.794	0.954
	Opacity	16.1	0.561	1.866
bartender (MPEG GSC)	Color (DC)	6.9	0.916	0.497
	Color (AC)	9.0	0.809	0.929
	Scale	7.0	0.960	0.273
	Rotation	10.1	0.762	1.071
	Opacity	23.9	0.492	2.067
cinema (MPEG GSC)	Color (DC)	7.3	0.906	0.553
	Color (AC)	9.8	0.814	0.931
	Scale	6.4	0.952	0.321
	Rotation	12.1	0.754	1.121
	Opacity	25.8	0.551	2.008

따라서 압축 효율을 향상시키기 위해서는 양자화 이후의 심볼 분포만을 고려하는 것이 아니라, 양자화 이전 실수 잠재 표현의 분포를 학습 과정에서 직접 제어할 필요가 있다. 실수 잠재 표현을 0 주변으로 집중시킬 수 있다면 양자화 이후 정수 잠재 표현의 0 심볼의 출현 비율을 증가시켜 엔트로피를 감소시킬 수 있으며, 이는 기존 디코더나 엔트로피 부호화 구조를 변경하지 않고도 압축 효율을 향상시킬 수 있음을 의미한다. 다음 절에서는 이를 위한 L1 기반 잠재 표현 희소화를 제안한다.

3. L1 기반 잠재 표현 희소화

잠재 분포와 압축 효율 분석을 통하여 양자화 이전 실수 잠재 표현의 분포를 제어하면 양자화 이후 정수 잠재 표현의 심볼 분포를 압축에 유리한 형태로 유도할 수 있음을 확인하였다. 따라서 본 논문에서는 학습 과정에서 실수 잠재 표현을 0 중심으로 집중시키기 위해 L1 기반 잠재 표현 희소화를 제안한다. 여기서 희소성 (Sparsity)은 신호의 대부분 성분이 0의 값을 갖는 성질을 의미하며, 본 논문에서는 정수 잠재 표현에서 0 심볼이 차지하는 비율 (zero-ratio)로 정량화한다. 희소성이 높을수록 심볼 분포가 0에 집중되어 엔트로피가 감소하므로 부호화에 필요한 평균 비트수가 줄어들며, 잠재 표현 희소화는 학습 과정에서 이러한 희소성을 높이도록 분포를 유도하는 것을 말한다. L1 정규화가 희소성을 유도하는 이유는 기울기 특성에 있다. L1 항의 기울기는 값의 크기와 무관하게 부호에 따라 일정하므로, 재구성 손실 감소에 대한 기여가 정규화 강도보다 작은 잠재 표현은 학습이 진행됨에 따라 0으로 수렴하고, 기여가 큰 잠재 표현은 일정한 편향만 받은 채 유지된다. 즉 L1 정규화는 압축 효율 대비 중요도가 낮은 성분을 선택적으로 0에 집중시키며, 그 결과 양자화 이후 0 심볼의 출현 비율이 증가하여 엔트로피 부호화 효율이 향상된다. 또한 변화가 작은 프레임 간 잔차를 0으로 표현하도록 유도하므로, 동적 장면의 시간적 중복성과도 자연스럽게 부합한다.

위치를 제외한 속성 집합을 $C=\{\mathrm{Color(DC)}, \mathrm{Color(AC)}, \mathrm{Scale}, \mathrm{Rotation}, \mathrm{Opacity}\}$라 할 때, 제안하는 잠재 표현의 희소화 손실은 다음과 같이 정의된다.

$$L_{sparse} = \sum_{c \in C} \frac{1}{N_c L_c} \sum_i \sum_j \left| \hat{l}_{c,i,j} \right| \tag{4}$$

여기서 $\hat{l}_{c,i,j}$는 속성 c의 i번째 가우시안에 대한 j번째 차원의 실수 잠재 표현이며, N_c와 L_c는 각각 해당 속성의 가우시안 수와 잠재 차원을 의미한다.

제안하는 희소화 손실은 기존 QUEEN의 목적 함수에 추가되어 전체 손실 함수는 다음과 같이 정의된다.

$$L_{propose} = L_{rec} + \lambda_{gate} L_{gate} + \lambda_{sparse} L_{sparse} \tag{5}$$

여기서 λ_{sparse}는 재구성 품질과 압축 효율 사이의 균형 (Trade-off)을 조절하는 가중치이다.

한편, L1 항은 0에서 미분이 불가능하므로 원칙적으로는 proximal gradient와 같은 별도의 최적화 기법이 고려될 수 있다. 본 방법에서는 자동 미분이 제공하는 subgradient를 사용하는 기존 QUEEN의 최적화 구성을 그대로 유지한다. 정확히 0인 해가 필요한 일반적인 희소 최적화 문제와 달리, 본 구조에서는 반올림 양자화가 0 주변 구간 전체를 0 심볼로 매핑하므로 잠재 표현이 정확히 0에 수렴할 필요가 없기 때문이다. 따라서 제안 방법은 최적화기나 학습률 등의 변경 없이 손실 항 추가만으로 동작한다.

제안하는 희소화는 위치를 제외한 속성 잠재 표현에만 적용한다. 위치 정보는 Hard Concrete Gate를 이용하여 별도로 희소화되므로 정규화 대상에서 제외하였다. 또한 첫 번째 프레임은 잔차가 아닌 기준 장면을 표현하므로 희소화 손실을 적용하지 않는다. 따라서 제안하는 방법은 QUEEN의 디코더, 게이트 및 엔트로피 부호화 구조를 변경하지 않고 손실 함수에 하나의 정규화 항만 추가하여 적용할 수 있다. 학습 결과로 디코더 파라미터와 심볼 확률 분포의 값은 달라지지만, 이는 기존 QUEEN에서도 프레임마다 전송되는 정보이므로 복호 절차와 비트스트림 구성에는 어떠한 변경도 발생하지 않는다. 즉 제안하는 방법은 복호기 관점의 규격 변화가 없는 비규범적 (Non-normative) 인코더 측 최적화에 해당한다. 한편, 잠재 표현을 수축시키는 대안적 정규화와의 비교는 4장 4절에서 정량적으로 다룬다.

Ⅳ. 실험 결과

본 장에서는 제안하는 L1 기반 잠재 표현 희소화의 효과를 다양한 관점에서 평가한다. 먼저 실험 환경을 설명한 후, 기존 QUEEN과의 율-왜곡 (Rate-Distortion) 성능을 비교하여 압축 성능을 평가한다. 이어서 잠재 표현의 심볼 분포 변화를 분석하여 제안하는 방법의 동작 원리를 검증한다. 또한 표준편차 기반 정규화와 비교하여 L1 정규화의 효과를 분석하고, 정성적 비교를 통해 재구성 품질을 평가한다. 마지막으로는 제안하는 방법의 계산 복잡도를 분석한다.

1. 실험 설정

제안 방법의 성능을 평가하기 위해 Neural 3D Video (N3DV)[24] 데이터셋의 5개 시퀀스(coffee_martini, cook_spinach, cut_roasted_beef, flame_steak, sear_steak)와 MPEG Gaussian Splat Coding (GSC)[25] 데이터셋의 2개 시퀀스 (bartender, cinema)를 사용하였다. N3DV는 2704×2028 해상도의 50개 프레임, MPEG GSC는 1920×1080 해상도의 32개 프레임으로 학습하였으며, 각 시퀀스에서 한 개의 시점을 평가용으로 분리하였다.

모든 평가 지표는 첫 프레임을 제외한 잔차 프레임 (N3DV 49개, MPEG GSC 31개)에 대해 평균을 취하여 제시하였다. 제안하는 L1 정규화가 작용하는 대상이 잔차 프레임이며, 첫 프레임은 잔차가 아닌 전체 장면을 표현하므로 정규화 대상에서 제외되기 때문이다. 재구성 품질은 PSNR과 LPIPS[26]로, 압축 효율은 프레임당 저장 용량 (MB)으로 측정하였다. 저장 용량은 양자화된 정수 잠재 표현의 엔트로피 부호화 비트량, 위치 잔차와 게이트 정보, 속성별 선형 디코더의 파라미터를 포함하며, 크기가 작은 심볼 확률 분포 등의 부가 정보는 제외하였다. 이는 기준 구조인 QUEEN의 저장 용량 산출 방식과 동일하다.

학습 설정은 제안하는 L1 정규화 항을 제외하고 기준 구조인 QUEEN[9]의 기본 구성을 그대로 사용하였다. 모든 실험은 단일 NVIDIA RTX A5000 (24GB)에서 Adam optimizer로 수행하였으며, 첫 프레임은 500 epoch, 이후 각 잔차 프레임은 10 epoch 학습하였다. 구면조화함수 (SH)의 차수는 2이고, 가우시안 위치 초기화에는 단안 깊이 추정[27]을 활용하였다. 재구성 손실은 픽셀 단위 L1 손실과 가중치 0.2의 D-SSIM 손실의 합으로 구성되며, 위치 잔차의 희소화를 위한 게이트 손실이 더해진다. 주요 하이퍼파라미터는 [표 2]에 정리하였으며, 제안하는 방법은 잔차 프레임에 적용되는 L1 정규화 가중치 λ_{sparse} 만을 추가로 도입한다. 본 4장의 모든 압축 성능 지표는 학습 과정의 무작위성에 의한 편차를 줄이기 위해 서로 다른 5개의 시드 (Random Seed)로 반복 학습한 결과를 평균하여 제시한다. 다만 속성별 정수 잠재 표현의 분포 분석과 4장 5절의 정성적 비교는 그 변화가 시드에 둔감하고 경향이 일관되므로, 대표 시드 하나의 결과로 제시한다.

2. 기준 구조 (QUEEN) 대비 율-왜곡 성능 분석

제안하는 방법이 기존 QUEEN 대비 재구성 품질을 유지하면서 저장 용량을 감소시키는지를 검증하기 위하여 율-

표 2. 주요 학습 하이퍼파라미터
Table 2. Key training hyperparameters

Category	Parameter	Value
Training	Optimizer	Adam
	Epochs (first / residual frame)	500 / 10
	SH degree	2
Loss	D-SSIM weight	0.2
	Gate L0 weight	0.01
Learning rate (residual)	Opacity / Rotation	0.05 / 0.05
	Scale / Color DC / Color AC	0.01 / 0.0025 / 0.000125
	Position / Latent decoder	0.0005 / $\{10^{-3}, 10^{-4}\}$
Latent dim	Color DC / AC / Scale / Rotation / Opacity	8 / 4 / 8 / 6 / 3
Proposed	L1 weight λ_{sparse}	{0, 10^{-5} ~ 5×10^{-2}}

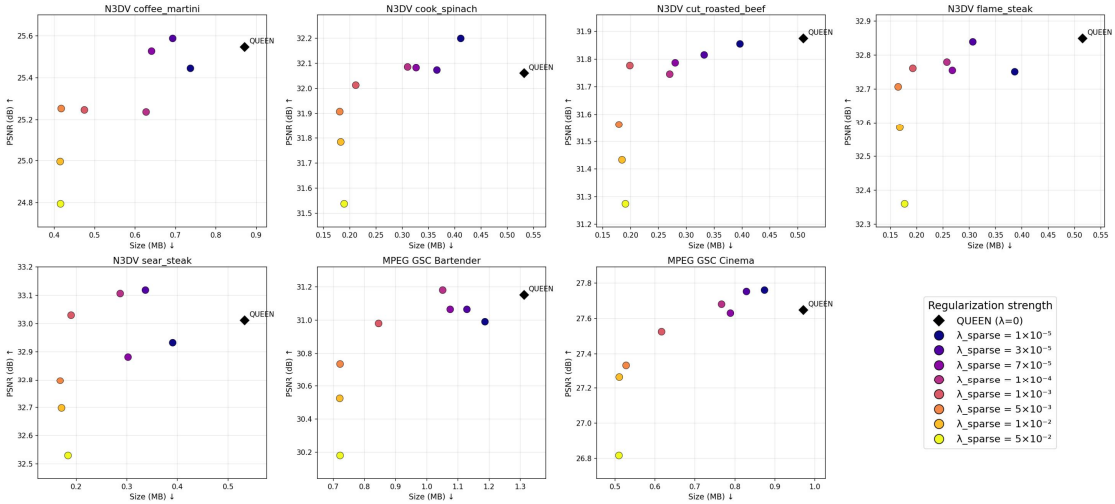

그림 2. 정규화 강도 λ_{sparse}에 따른 저장 용량-PSNR 산점도
Fig. 2. Storage-PSNR scatter for varying λ_{sparse}

왜곡 성능을 비교하였다. 비교는 7개 시퀀스에 대해 다양한 정규화 강도 (λ_{sparse})를 적용하여 수행하였으며, PSNR과 저장 용량 사이의 Trade-off를 분석하였다.

[그림 2]는 정규화 강도 λ_{sparse}에 따른 저장 용량과 PSNR의 관계를 산점도로 나타낸 것이다. 기준 구조 (QUEEN, $\lambda_{sparse}=0$)가 단일 동작점만 제공하는 것과 달리, 제안하는 방법은 λ_{sparse} 조절을 통해 압축률과 화질 사이의 Trade-off를 폭넓게 조절할 수 있다. 7개 시퀀스 모두에서 λ_{sparse}를 증가시키면 저장 용량이 크게 감소하는 한편, 적절한 강도 구간에서는 PSNR이 기준 구조 수준 부근에 머무르다가 특정 강도를 넘어서면 화질이 급격히 저하되는 공통된 경향이 관찰된다.

이러한 경향에서, λ_{sparse}가 대략 1×10^{-3} 부근인 구간이 저장 용량 감소와 화질 유지 사이의 균형 측면에서 가장 효율적인 것으로 판단된다.

[표 3]은 화질 저하를 최소화하는 동작점 (PSNR 손실 0.3 dB 이내)에서의 저장 용량 절감을 정량화한 것으로, 7개 시퀀스에서 PSNR 손실 0.3 dB 이내로 평균 55.4%, 최대 68.4% (sear_steak)의 저장 용량이 절감되었다. 이때 LPIPS의 변화 또한 해당 동작점에서의 모든 시퀀스에서 +0.01 이내로 미미하며, 지각적 품질도 함께 유지됨을 확인하였다.

이러한 결과는 N3DV와 MPEG GSC 두 데이터셋에서 일관되게 관찰되었다. 절감률이 상대적으로 낮은 bartender와 cinema에서도 0.2 dB 이내의 화질 변화로 35% 이상의 저장 용량을 절감하였다. 즉 제안하는 L1 기반 잠재 표현

표 3. 화질 저하를 최소화하는 동작점에서의 저장 용량 절감
Table 3. Compression at the operating point (minimal PSNR loss)

Sequence	Baseline PSNR (dB) / Size (MB)	λ_{sparse}	L1 PSNR (dB) / Size (MB)	△PSNR ↑	△LPIPS ↓	Reduction ↑
coffee_martini	25.55 / 0.871	5×10^{-3}	25.25 / 0.417	-0.29	+0.010	52.1%
cook_spinach	32.06 / 0.532	5×10^{-3}	31.91 / 0.180	-0.15	+0.004	66.1%
cut_roasted_beef	31.88 / 0.510	1×10^{-3}	31.78 / 0.198	-0.10	+0.003	61.1%
flame_steak	32.85 / 0.515	5×10^{-3}	32.71 / 0.165	-0.14	+0.003	68.1%
sear_steak	33.01 / 0.533	5×10^{-3}	32.80 / 0.168	-0.21	+0.003	68.4%
bartender	31.15 / 1.311	1×10^{-3}	30.98 / 0.845	-0.17	+0.004	35.6%
cinema	27.65 / 0.971	1×10^{-3}	27.53 / 0.616	-0.12	+0.005	36.6%
Average						55.4%

희소화가 기존 압축 구조를 변경하지 않고도 저장 용량을 감소시키면서 재구성 품질을 유지할 수 있음을 보여준다.

3. L1 정규화에 의한 잠재 표현 분포 변화 분석

3장에서 분석한 바와 같이 제안하는 방법은 양자화 이전 잠재 표현을 0 중심으로 집중시켜 양자화 이후 심볼 분포를 압축에 유리한 형태로 유도하는 것을 목표로 한다. 본 절에서는 실제 학습된 정수 잠재 표현의 심볼 분포가 어떻게 변화하는지를 분석하여 제안한 방법의 동작 원리를 검증한다.

[표 4]는 [표 1]과 동일한 네 시퀀스에 대해 L1 정규화 적용 전후의 속성별 심볼 분포를 비교한 것이다. 모든 속성에서 zero-ratio가 증가하고 심볼당 비트 수가 감소하였으며, 특히 [표 1]에서 비트 소모가 가장 컸던 불투명도와 회전 속성에서 그 변화가 두드러진다. 예를 들어 coffee_martini 시퀀스의 불투명도 속성은 zero-ratio가 0.610에서 1.000으로 증가하여 심볼당 비트가 1.721에서 0으로 감소하였고, 회전 속성 역시 zero-ratio가 0.828에서 0.996으로 증가하였다. 즉, 가장 많은 저장 용량을 소모하던 속성들이 L1 정규화에 의해 정수 잠재 표현이 0 주변으로 집중되어 비트 수가 크게 감소한다.

[그림 3-a, 3-b]는 이러한 변화를 정수 잠재 표현의 심볼 분포로 시각화한 것이다. 기준 구조에서 불투명도와 회전 속성은 0을 정점으로 하되 넓은 범위의 심볼값에 걸쳐 분포하지만, λ_{sparse} 가 증가함에 따라 분포가 0으로 집중되며 충분한 강도에서는 대부분의 잔차가 0이 된다. 반면 이미 0 부근에 집중되어 있던 스케일 속성은 거의 변화하지 않는데, 이는 L1 정규화가 정수 잠재 표현의 분포가 넓게 퍼져 압축에 불리한 속성을 선택적으로 희소화함을 보여준다. 이러한 경향은 [그림 3-a, 3-b]의 두 시퀀스 (cinema, coffee_martini)를 비롯해 [표 4]의 모든 시퀀스에서 일관되게 나타난다.

표 4. L1 정규화 적용 전후의 속성별 심볼 분포
(표 1과 동일한 네 시퀀스에 대해 QUEEN 기준 구조와 L1 정규화 적용 후의 속성별 zero-ratio와 심볼당 비트를 비교하였으며, 대표 시드 하나의 결과이다. L1 정규화 강도는 표 3의 동작점 (coffee_martini·sear_steak $\lambda_{sparse} = 5 \times 10^{-3}$, bartender·cinema 1×10^{-3})을 사용하였다.)
Table 4. Per-attribute symbol distribution before and after L1 regularization
(For the same four sequences as Table 1, the per-attribute zero-ratio and bits-per-symbol of the QUEEN baseline and the L1-regularized model are compared, using a single representative seed. The L1 regularization strengths are the operating points from Table 3 (coffee_martini·sear_steak $\lambda_{sparse} = 5 \times 10^{-3}$, bartender·cinema 1×10^{-3}).)

Dataset	Attribute	zero-ratio (Base)	zero-ratio (L1)	bit/symbol (Base)	bit/symbol (L1)
coffee_martini (N3DV)	Color (DC)	0.961	0.998	0.278	0.026
	Color (AC)	0.850	0.927	0.797	0.471
	Scale	0.981	0.996	0.151	0.038
	Rotation	0.828	0.996	0.840	0.043
	Opacity	0.610	1.000	1.721	0.000
sear_steak (N3DV)	Color (DC)	0.959	0.998	0.286	0.019
	Color (AC)	0.847	0.934	0.814	0.429
	Scale	0.980	0.998	0.158	0.024
	Rotation	0.794	0.998	0.954	0.028
	Opacity	0.561	0.999	1.866	0.008
bartender (MPEG GSC)	Color (DC)	0.916	0.969	0.497	0.231
	Color (AC)	0.809	0.904	0.929	0.559
	Scale	0.960	0.983	0.273	0.144
	Rotation	0.762	0.967	1.071	0.247
	Opacity	0.492	0.942	2.067	0.426
cinema (MPEG GSC)	Color (DC)	0.906	0.961	0.553	0.280
	Color (AC)	0.814	0.890	0.931	0.639
	Scale	0.952	0.974	0.321	0.197
	Rotation	0.754	0.957	1.121	0.311
	Opacity	0.551	0.967	2.008	0.264

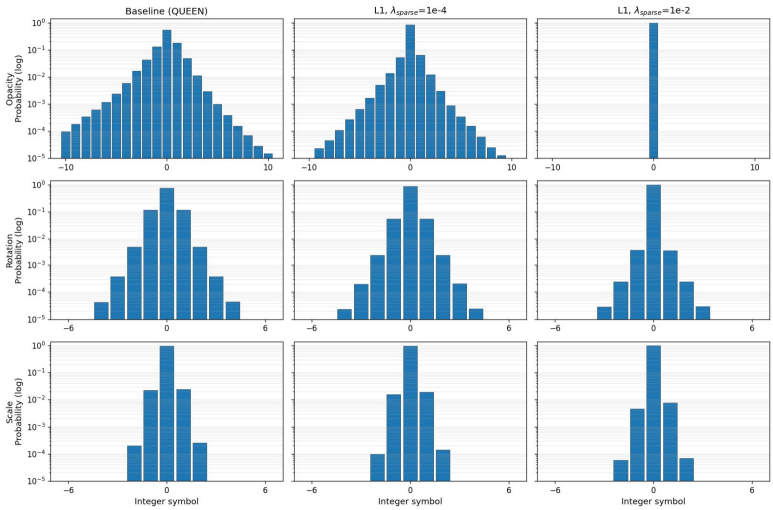

그림 3-a. cinema 시퀀스의 속성별 (불투명도·회전·스케일) 정수 잠재 표현의 심볼 분포 (좌측부터 기준 구조 (QUEEN), 제안하는 L1 정규화 ($\lambda_{sparse} = 10^{-4}, 10^{-2}$)의 결과이며, 각 그래프의 가로축은 정수 심볼값, 세로축은 로그 스케일로 나타낸 출현 확률이다.)

Fig. 3-a. Integer latent symbol distributions (opacity, rotation, scale) for the cinema sequence.

(From left: baseline(QUEEN) and the proposed L1 regularization with $\lambda_{sparse} = 10^{-4}$ and 10^{-2}. The horizontal axis denotes the integer symbol value and the vertical axis the occurrence probability in log scale.)

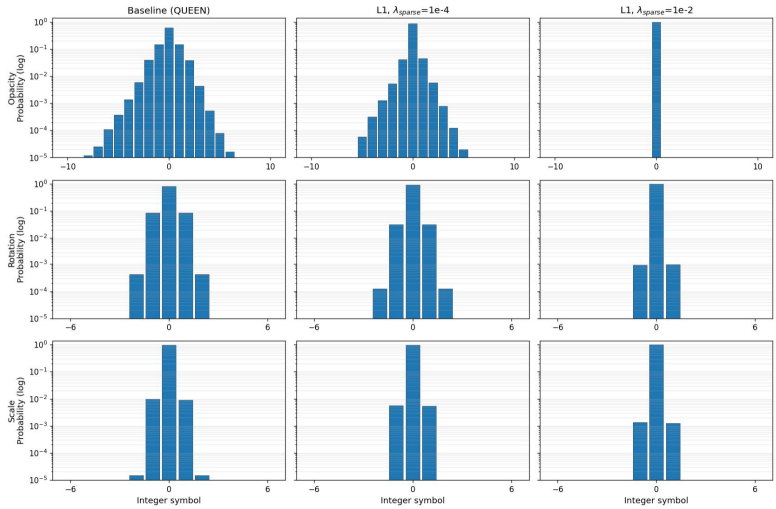

그림 3-b. coffee_martini 시퀀스에 대한 동일한 비교

Fig. 3-b. The same comparison for the coffee_martini sequence

한편, 이러한 압축은 가우시안의 개수를 줄이는 것이 아니라 각 가우시안의 잠재 표현 비트 수를 감소시킴으로써 이루어진다. [표 5]에서 볼 수 있듯이 제안하는 방법 적용 전후로 가우시안의 개수는 약 1.4% 이내로 거의 변하지 않았으며, 이는 제안하는 방법이 위치 게이트 기반의 프루닝과 독립적으로 양자화 대상 속성 잠재 표현의 분포를 직접 제어하여, 가우시안 개수가 아닌 잠재 표현의 엔트로피 감소를 통해 압축을 달성함을 의미한다.

표 5. L1 정규화 적용 전후의 가우시안 개수 비교
(표 3의 동작점 (coffee_martini·sear_steak $\lambda_{sparse} = 5 \times 10^{-3}$, bartender·cinema 1×10^{-3})에서 학습한 대표 시드 하나의 결과이며, 잔차 프레임에 대해 평균한 가우시안 개수이다.)
Table 5. Number of Gaussians before and after L1 regularization
(The number of Gaussians averaged over residual frames, using a single representative seed trained at the operating points (coffee_martini·sear_steak $\lambda_{sparse} = 5 \times 10^{-3}$, bartender·cinema 1×10^{-3}) from Table 3.)

Sequence	# Gaussians (Base)	# Gaussians (L1)	Change
coffee_martini	331,045	331,944	+0.27%
sear_steak	222,723	219,761	-1.33%
bartender	378,106	378,903	+0.21%
cinema	255,333	256,312	+0.38%

4. 표준편차 기반 정규화와의 비교

본 논문에서는 잠재 표현을 0 중심으로 수축시키기 위한 방법으로 L1 기반 정규화를 제안하였다. 한편, 잠재 표현을 집중시키는 방법으로는 분포의 표준편차를 감소시키는 표준편차 기반 정규화도 고려할 수 있으며, 이는 QUEEN[9]의 부록에서도 선택적 개선 방안으로 언급된 바 있다. 두 방법 모두 잠재 표현의 분포를 수축시키는 효과를 가지지만, 분포를 변화시키는 방식에는 차이가 존재한다. 따라서 본 절에서는 동일한 학습 환경에서 표준편차 기반 정규화와 L1 기반 정규화를 비교하여 제안한 방법의 효과를 검증한다.

표준편차 기반 정규화는 각 속성 잠재 표현의 표준편차를 감소시키도록 학습되므로, 분포 전체를 평균을 중심으로 비례적으로 수축시키는 특성을 가진다. 반면 L1 정규화는 각 잠재 표현의 절댓값에 직접 작용하여 크기가 작은 잠재 표현을 우선적으로 0으로 유도한다. 따라서 두 방법은 잠재 분포를 변화시키는 방식이 서로 다르며, 이러한 차이는 양자화 이후 정수 잠재 표현의 심볼 분포와 엔트로피 부호화 효율에도 영향을 미칠 것으로 기대된다.

이를 검증하기 위하여 동일한 기준 구조와 학습 조건에서 표준편차 기반 정규화와 제안하는 L1 기반 정규화를 각각 적용하였다. 두 정규화 항은 손실 함수에 기여하는 값의

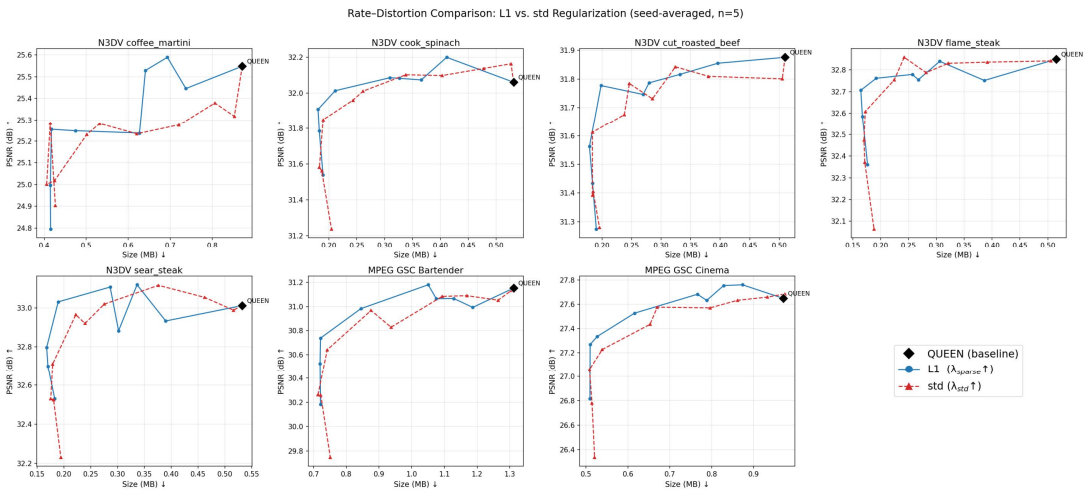

그림 4. L1 기반 정규화와 표준편차 기반 정규화의 시드 평균 율-왜곡 곡선 (7개 시퀀스, 시드 5회 평균)
(각 곡선은 정규화 강도 순으로 연결하였으며, $\lambda_{sparse} = \lambda_{std} = 0$인 기준 구조 (QUEEN)를 공유한다.)
Fig. 4. Seed-averaged rate-distortion curves of the proposed L1-based and the standard-deviation-based regularization for seven sequences (averaged over 5 seeds)
(Each curve connects operating points in order of regularization strength and shares the baseline (QUEEN).)

크기 (Scale)가 서로 다르므로, 동일한 가중치 값으로는 같은 세기의 정규화 효과를 얻을 수 없다. 따라서 각 방법의 정규화 가중치를 방법별로 독립적인 범위 (L1 기반 정규화 가중치 $\lambda_{sparse}=1\times10^{-5}\sim5\times10^{-2}$, 표준편차 기반 정규화 가중치 $\lambda_{std}=1\times10^{-8}\sim5\times10^{-2}$)에서 변화시키며 학습을 반복하여, 두 방법이 기준 구조 부근의 화질부터 강한 압축까지 유사한 저장 용량-PSNR 영역을 포괄하는 율-왜곡 곡선을 각각 구성하였다. 각 동작점은 시드 5회 평균이며, 두 곡선은 정규화를 적용하지 않은 기준 구조 (QUEEN)를 공통 시작점으로 공유한다. 이때 비교는 가중치 값이 아닌 율-왜곡 공간에서 수행되므로 두 방법의 가중치가 일대일로 대응될 필요는 없다.

[그림 4]는 두 정규화 방법의 율-왜곡 곡선을 비교한 결과이며, [표 6]은 각 시퀀스에서 계산한 BD-rate를 나타낸다. BD-rate는 두 율-왜곡 곡선으로부터 동일 화질에서의 평균 저장 용량 차이를 산출한 지표로, 음수일수록 제안하는 방법이 더 적은 저장 용량을 사용함을 의미한다. 구체적으로는 각 정규화 방법의 모든 동작점 (시드 5회 평균의 저장 용량과 PSNR)에 대해 저장 용량의 로그값을 PSNR의 3차 다항식으로 최소 제곱 적합하고, 두 곡선의 PSNR이 겹치는 구간에서 적분한 평균 로그 용량의 차이를 백분율로 환산하였다[28].

[그림 4]에서 확인할 수 있듯이 두 정규화 모두 기준 구조 대비 저장 용량을 감소시키는 효과를 보였으나, 대부분의 시퀀스에서 L1 기반 정규화가 표준편차 기반 정규화보다 더 우수한 율-왜곡 특성을 보였다. 또한 [표 6]에서 제안하는 방법의 BD-rate는 7개 시퀀스 전체에서 음수를 기록하였으며 평균 약 -7.4%로 표준편차 기반 정규화보다 일관되게 높은 압축 효율을 달성하였다. 이와 같이 잠재 표현의 분포를 단순히 수축시키는 것보다 0 중심의 희소성을 직접 유도하는 것이 압축 효율 향상에 더욱 효과적임을 보여준다. 따라서 본 논문에서는 잠재 표현 희소화를 위한 정규화 방법으로 L1 기반 정규화를 채택하였다.

표 6. L1 기반 정규화와 표준편차 기반 정규화의 BD-rate 비교
(시드 5회 평균, 음수 = L1이 동일 화질에서 더 적은 저장 용량)
Table 6. BD-rate comparison between the proposed L1-based and the standard-deviation based regularization
(Averaged over 5 seeds; a negative value indicates smaller storage for L1 at the same quality.)

Sequence	BD-rate↓
coffee_martini	-10.0%
cook_spinach	-11.0%
cut_roasted_beef	-7.1%
flame_steak	-1.7%
sear_steak	-8.0%
bartender	-4.6%
cinema	-9.7%
Average	-7.4%

5. 정성적 분석

앞선 정량적 평가에서 제안하는 방법이 기존 QUEEN 대비 재구성 품질 저하를 최소화하면서 저장 용량을 효과적으로 감소시킴을 확인하였다. 그러나 PSNR 및 LPIPS와 같은 평균 지표만으로는 특정 영역에서 발생할 수 있는 국소적인 압축 왜곡까지 충분히 평가하기 어렵다. 따라서 본 절에서는 실제 렌더링 결과를 비교하여 제안하는 방법이 저장 용량을 크게 감소시키면서도 기존 QUEEN과 동등한 시각적 품질을 유지하는지를 정성적으로 분석한다.

[그림 5]는 MPEG GSC의 bartender와 N3DV의 sear_steak 시퀀스에 대해 원본(Ground-Truth), QUEEN 기준 구조, 제안하는 방법 (L1)의 테스트 시점 렌더링 결과를 비교한 것이다. 제안하는 방법은 [표 3]의 동작점 (bartender $\lambda_{sparse}=1\times10^{-3}$, sear_steak $\lambda_{sparse}=5\times10^{-3}$)을 사용하였으며, 잔차 프레임 평균으로는 각각 약 36%, 68%의 저장 용량을 절감한다. 각 결과의 상단에는 해당 프레임의 저장 용량, PSNR 및 LPIPS를 함께 표기하였으며, 압축 왜곡이 두드러질 가능성이 높은 큰 움직임 영역 (빨강), 반사광 영역 (파랑), 고주파 텍스처 영역 (초록)을 확대하여 비교하였다.

확대 영역에서 확인할 수 있듯이, 제안하는 방법은 저장 용량을 각각 약 36%와 68% 감소시켰음에도 QUEEN 기준 구조와 비교하여 시각적으로 구별되는 추가적 화질 저하를 보이지 않는다. 특히 움직임이 큰 인물의 손과 팔 영역에서는 경계가 안정적으로 유지되며, 반사광이 포함된 금속 및 유리 표면에서도 추가적인 왜곡이 관찰되지 않는다. 또한 세밀한 텍스처 영역에서도 질감과 구조가 기준 구조와 거의 동일하게 유지되어 고주파 성분의 손실이 두드러지지 않는다.

그림 5. 테스트 시점 렌더링 결과의 정성적 비교
(좌측부터 원본 (Ground-Truth), QUEEN 기준 구조, 제안하는 방법 (L1)이며, 기준 구조와 제안 방법의 상단에 해당 프레임의 저장 용량 / PSNR / LPIPS를 표기하였다. 색 상자는 확대 영역 (빨강: 움직임, 파랑: 반사광, 초록: 텍스처)을 나타낸다. 상단부터 bartender 시퀀스 ($\lambda_{sparse} = 1 \times 10^{-3}$), sear_steak 시퀀스($\lambda_{sparse} = 5 \times 10^{-3}$)이다.)
Fig. 5. Qualitative comparison of rendered test views
(From left to right: Ground-Truth, QUEEN baseline, and the proposed method (L1). The storage size / PSNR / LPIPS of each frame is annotated at the top of the baseline and proposed-method results. Colored boxes denote the zoomed regions (red: motion, blue: specular, green: texture). The top and bottom rows correspond to the bartender ($\lambda_{sparse} = 1 \times 10^{-3}$) and sear_steak ($\lambda_{sparse} = 5 \times 10^{-3}$) sequences, respectively.)

원본 영상과 기준 구조 사이에는 기존 QUEEN의 압축 과정에서 발생하는 재구성 오차가 존재하지만, 이는 기준 구조 자체의 특성에 기인한 것이다. 반면 제안하는 방법은 기준 구조 대비 저장 용량을 크게 감소시키면서도 이러한 재구성 특성을 그대로 유지하며, 추가적인 시각적 열화를 거의 발생시키지 않는다. 이는 제안한 L1 기반 잠재 표현 희소화가 잠재 표현의 분포만을 제어하여 엔트로피를 감소시키고, 기존 압축 구조의 재구성 능력은 유지함을 의미한다. 따라서 제안하는 방법은 압축 효율과 시각적 품질 보존을 동시에 달성하며, 이는 앞선 정량적 평가와 잠재 분포 분석 결과를 뒷받침한다.

6. 계산 복잡도 분석

제안하는 방법은 손실 함수에 정규화 항 하나만 추가하므로 별도의 학습 파라미터나 구조 변경을 필요로 하지 않으며, 추가 연산은 잠재 표현의 절댓값 평균을 계산하는 비용에 한정된다. [표 7]은 [표 3]의 동작점에서 잔차 프레임 한 장을 학습하는 데 소요된 평균 시간을 비교한 것이다. 7개 시퀀스 평균 학습 시간은 기준 구조 7.38초, 제안하는 방법 7.25초로 학습 시간의 증가는 관측되지 않았다. 또한 제안하는 방법은 복호 절차를 변경하지 않으며 [표 5]에서 볼 수 있듯 가우시안의 개수 또한 거의 변화시키지 않으므로, 복호화 및 렌더링 복잡도에도 영향을 미치지 않는다.

표 7. L1 정규화 적용 전후의 학습 시간 비교
(표 3의 동작점에서 학습한 대표 시드 하나의 결과이며, 잔차 프레임 한 장당 평균 학습 시간(초)이다.)
Table 7. Training time before and after L1 regularization
(The training time per residual frame in seconds, using a single representative seed trained at the operating points from Table 3.)

Sequence	Training time (Base)	Training time (L1)	Change
coffee_martini	7.67	7.57	-1.3%
cook_spinach	7.69	7.64	-0.7%
cut_roasted_beef	7.16	6.91	-3.5%
flame_steak	7.54	7.35	-2.5%
sear_steak	7.61	7.46	-2.0%
bartender	7.72	7.51	-2.7%
cinema	6.29	6.29	0.0%
Average	7.38	7.25	-1.8%

V. 결 론

본 논문에서는 동적 3D 가우시안 스플래팅 기반 자유시점 비디오 압축 방법인 QUEEN에서 양자화 이전 잔차 잠재 표현의 분포를 직접 제어하기 위한 L1 기반 잠재 표현 희소화를 적용하고, 그 효과를 실험적으로 분석하였다. 제안하는 방법은 위치를 제외한 속성의 잔차 잠재 표현에 L1 정규화를 적용하여 양자화 이후 0 심볼의 출현 비율을 증가시키고, 이를 통해 엔트로피 부호화 효율을 향상시킨다. 또한 기존 QUEEN 인코더, 디코더, 양자화 및 엔트로피 부호화 구조를 변경하지 않고 손실 함수에 하나의 정규화 항만 추가하므로, 추론 구조와 복호화 복잡도를 증가시키지 않는다는 장점을 가진다.

N3DV 5개 시퀀스와 MPEG GSC 2개 시퀀스에 대한 실험에서, 제안 방법은 화질 저하가 거의 없는 동작점 (PSNR 손실 0.3 dB 이내)에서 QUEEN 대비 평균 55.4%, 최대 68.4%의 잔차 저장 용량을 절감하였다. 또한 LPIPS와 정성적 비교 결과에서도 기준 구조와 유사한 시각적 품질을 유지하면서도 저장 용량을 효과적으로 감소시킴을 확인하였다.

잠재 표현 분포 분석을 통해 제안한 L1 정규화가 비트 예산이 큰 불투명도와 회전 속성의 잔차를 선택적으로 0에 수렴시키는 반면, 이미 0 심볼의 비율이 높은 스케일 속성에는 거의 영향을 미치지 않음을 확인하였다. 또한 가우시안 개수는 정규화 전후 거의 변화하지 않아 (약 1.4% 이내) 저장 용량 감소가 가우시안 수 감소가 아니라 잔차 잠재 표현의 희소화와 이에 따른 엔트로피 감소에서 비롯됨을 확인하였다.

또한 표준편차 기반 정규화와 비교를 통해 제안하는 L1 기반 희소화가 동일 화질 조건에서 평균 7.4%의 추가적인 BD-rate 개선을 달성하여 잠재 표현을 압축에 유리한 형태로 유도하는 데 더욱 효과적임을 검증하였다. 아울러 잔차 프레임당 평균 학습 시간은 기준 구조 7.38초, 제안하는 방법 7.25초로 학습 시간의 증가가 관측되지 않아, 추가적인 계산 비용 없이 압축 효율을 향상시킬 수 있음을 확인하였다.

본 연구는 동적 가우시안 스플래팅 기반 자유시점 비디오 압축에서 잠재 표현 자체의 표현 능력이 아닌, 잠재 표현의 분포를 제어하는 접근의 효과를 체계적으로 분석하고 실증하였다는 점에서 의의를 갖는다. 제안하는 정규화는 특정 압축 구조에 종속되지 않으므로, 별도의 엔트로피 모델 없이 잠재 표현을 양자화·부호화하는 다양한 가우시안 스플래팅 압축 기법에도 적용 가능할 것으로 기대된다.

한편, 제안하는 정규화는 프레임 간 잔차 잠재 표현을 대상으로 하므로, 잔차가 정의되지 않는 첫 프레임에는 적용되지 않는다. 향후에는 시퀀스 특성에 따라 정규화 강도를 자동으로 결정하는 적응형 정규화 기법과 첫 프레임 압축을 위한 보다 효율적인 장면 표현 기법을 결합하여 비디오 전체의 압축 효율을 향상시키는 연구를 수행할 계획이다.

참 고 문 헌 (References)

[1] A. Smolic, “3D video and free viewpoint video－From capture to display,” Pattern Recognition, Vol.44, No.9, pp.1958-1968, 2011.

doi: https://doi.org/10.1016/j.patcog.2010.09.005

[2] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, and L. Yu, "MPEG Immersive Video Coding Standard," Proceedings of the IEEE, Vol.109, No.9, pp.1521-1536, 2021.
doi: https://doi.org/10.1109/JPROC.2021.3062590

[3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," Proceedings of European Conference on Computer Vision (ECCV), Glasgow, UK, pp.405-421, 2020.
doi: https://doi.org/10.1007/978-3-030-58452-8_24

[4] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," ACM Transactions on Graphics, Vol.42, No.4, Article 139, July 2023.
doi: https://doi.org/10.1145/3592433

[5] G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian, and X. Wang, "4D Gaussian splatting for real-time dynamic scene rendering," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp.20310-20320, 2024.
doi: https://doi.org/10.1109/CVPR52733.2024.01920

[6] J. C. Lee, D. Rho, X. Sun, J. H. Ko, and E. Park, "Compact 3D Gaussian representation for radiance field," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp.21719-21728, 2024.
doi: https://doi.org/10.1109/CVPR52733.2024.02052

[7] S. Girish, K. Gupta, and A. Shrivastava, "EAGLES: Efficient accelerated 3D Gaussians with lightweight encodings," Proceedings of European Conference on Computer Vision (ECCV), Milan, Italy, 2024.
doi: https://doi.org/10.1007/978-3-031-73036-8_4

[8] H. Lee, J. Byeon, M. Kim, and D. Sim, "4D Gaussian splatting compression framework compatible with commercial video codecs," Journal of Broadcast Engineering, Vol.31, No.1, pp.106-122, 2026.
doi: https://doi.org/10.5909/JBE.2026.31.1.106

[9] S. Girish, T. Li, A. Mazumdar, A. Shrivastava, D. Luebke, and S. De Mello, "QUEEN: Quantized efficient encoding of dynamic Gaussians for streaming free-viewpoint videos," Advances in Neural Information Processing Systems (NeurIPS), Vol.37, Vancouver, Canada, 2024.
doi: https://doi.org/10.52202/079017-1376

[10] Y.-T. Zhan, C.-Y. Ho, H. Yang, Y.-H. Chen, J. C. Chiang, Y.-L. Liu, and W.-H. Peng, "CAT-3DGS: A context-adaptive triplane approach to rate-distortion-optimized 3DGS compression," Proceedings of International Conference on Learning Representations (ICLR), Singapore, 2025.
doi: https://doi.org/10.48550/arXiv.2503.00357

[11] J. Sun, H. Jiao, G. Li, Z. Zhang, L. Zhao, and W. Xing, "3DGStream: On-the-fly training of 3D Gaussians for efficient streaming of photo-realistic free-viewpoint videos," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp.20675-20685, 2024.
doi: https://doi.org/10.1109/CVPR52733.2024.01954

[12] J. Chen, Q. Mao, Y. Bao, X. Meng, F. Meng, R. Wang, and Y. Liang, "Motion matters: Compact Gaussian streaming for free-viewpoint video reconstruction," Advances in Neural Information Processing Systems (NeurIPS), Vol.38, pp.133458-133482, San Diego, CA, USA, 2025.
doi: https://doi.org/10.52202/085713-4018

[13] J. Fu, Q. Gao, C. Wen, Y. Wu, S. Ma, J. Zhang, and J. Zhang, "ReCon-GS: Continuum-preserved Gaussian streaming for fast and compact reconstruction of dynamic scenes," Advances in Neural Information Processing Systems (NeurIPS), Vol.38, pp.149397-149424, San Diego, CA, USA, 2025.
doi: https://doi.org/10.52202/085713-4496

[14] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol.13, No.7, pp.560-576, 2003.
doi: https://doi.org/10.1109/TCSVT.2003.815165

[15] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol.22, No.12, pp.1649-1668, 2012.
doi: https://doi.org/10.1109/TCSVT.2012.2221191

[16] V. K. Goyal, "Theoretical foundations of transform coding," IEEE Signal Processing Magazine, Vol.18, No.5, pp.9-21, 2001.
doi: https://doi.org/10.1109/79.952802

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Series B, Vol.58, No.1, pp.267-288, 1996.
doi: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[18] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," Proceedings of International Conference on Learning Representations (ICLR), Toulon, France, 2017.
doi: https://doi.org/10.48550/arXiv.1611.01704

[19] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," Proceedings of International Conference on Learning Representations (ICLR), Vancouver, Canada, 2018.
doi: https://doi.org/10.48550/arXiv.1802.01436

[20] D. Minnen, J. Ballé, and G. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," Advances in Neural Information Processing Systems (NeurIPS), Vol.31, Montréal, Canada, 2018.
doi: https://doi.org/10.48550/arXiv.1809.02736

[21] L. Tang, J. Yang, R. Peng, Y. Zhai, S. Shen, and R. Wang, "Compressing streamable free-viewpoint videos to 0.1 MB per frame," Proceedings of the AAAI Conference on Artificial Intelligence, Vol.39, No.7, pp.7257-7265, 2025.
doi: https://doi.org/10.1609/aaai.v39i7.32780

[22] Q. Hu, Z. Zheng, H. Zhong, S. Fu, L. Song, X. Zhang, G. Zhai, and Y. Wang, "4DGC: Rate-aware 4D Gaussian compression for efficient streamable free-viewpoint video," Proceedings of IEEE/CVF

Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp.875-885, 2025.
doi: https://doi.org/10.1109/CVPR52734.2025.00090

[23] C. Louizos, M. Welling, and D. P. Kingma, “Learning sparse neural networks through L0 regularization,” Proceedings of International Conference on Learning Representations (ICLR), Vancouver, Canada, 2018.
doi: https://doi.org/10.48550/arXiv.1712.01312

[24] T. Li, M. Slavcheva, M. Zollhöfer, S. Green, C. Lassner, C. Kim, T. Schmidt, S. Lovegrove, M. Goesele, R. Newcombe, and Z. Lv, “Neural 3D video synthesis from multi-view video,” Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp.5521-5531, 2022.
doi: https://doi.org/10.1109/CVPR52688.2022.00544

[25] ISO/IEC JTC 1/SC 29/WG 7, Common Test Conditions for Gaussian Splat Coding, ISO/IEC JTC 1/SC 29/WG 7 N1414, Online, Jan. 2026.

[26] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, pp.586-595, 2018.
doi: https://doi.org/10.1109/CVPR.2018.00068

[27] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, “Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.44, No.3, pp.1623-1637, 2022.
doi: https://doi.org/10.1109/TPAMI.2020.3019967

[28] G. Bjøntegaard, “Calculation of average PSNR differences between RD-curves,” document VCEG-M33, ITU-T SG16/Q6, Austin, TX, USA, April 2001.

저 자 소 개

김 태 유

- 2025년 2월 : 광운대학교 컴퓨터공학과 학사
- 2025년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 석사과정
- ORCID : https://orcid.org/0009-0004-8165-8521
- 주관심분야 : 3D데이터압축, 영상압축, 컴퓨터비전

박 장 수

- 2016년 8월 : 광운대학교 컴퓨터공학과 학사
- 2019년 2월 : 광운대학교 컴퓨터공학과 석사
- 2019년 8월 ~ 2023년 1월 : 아이브스 주식회사 선임연구원
- 2023년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0009-0005-9312-6781
- 주관심분야 : 영상신호처리, 고성능컴퓨팅

심 동 규

- 1993년 2월 : 서강대학교 전자공학과 공학학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비젼 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : https://orcid.org/0000-0002-2794-9932
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전