

3D 최신 기술과 현황

□ 박성준, 조수빈, 서정일 / 동아대학교

요약

Neural Radiance Field의 계산 복잡도와 느린 렌더링 속도를 극복하고자 등장한 3D Gaussian Splatting의 기술 발전 흐름을 중심으로, 최신 3차원 장면 재구성 기법을 고찰한다. 3DGS는 이산적인 가우시안 입자 표현을 통해 고속성과 고품질 시각화를 달성하며, 실시간 응용에 적합한 방식으로 주목받고 있다. 이후 제안된 네 가지 확장인 단일 이미지 기반 재구성(AGG, SuGaR), 동적 장면 처리(4D-GS), 고정밀·고속 렌더링(3DLS), 효율적 학습 전략(Group Training)을 중심으로, 각 기술의 구조와 성능 개선 방안을 분석한다. 이를 통해 3DGS의 응용 가능성과 기술적 진화를 정리하고, 향후 확장 방향을 제시한다.

I. 서론

3차원 장면 재구성은 컴퓨터 비전과 그래픽스 분야에서 오랜 기간 핵심 연구과제로 자리잡아 왔으며, 최근 메타버스, 증강현실(AR)/가상현실(VR), 디지털 휴먼, 스마트 시티 등 실감형 미디어 기술의 발전과 함께 중요성이 더욱 강조되고 있다. 전통적인 3D 재구성 기술은 Multi-View Stereo(MVS), Structure-from-Motion(SfM), Simultaneous Localization and Mapping(SLAM)과 같은 다중 이미지 기반 기하학적 추정 기법을 통해 정밀한 3D 정보를 복원해 왔다[1, 2, 3]. 그러나 이러한 접근 방식은 복수의 정렬된 이미지를 요구하고, 복잡한 후처리 과정을 수반하기 때문에 실시간 처리나 확장성 면에서 한계

를 드러낸다.

딥러닝 기반 데이터 중심적 접근 방식의 도입은 이러한 한계를 극복하는 새로운 가능성을 열었다. 특히 Neural Radiance Field(NeRF)와 같은 신경 렌더링 기술은 고품질의 뷰를 구현하며 3D 재구성의 패러다임을 혁신적으로 전환시켰다[4]. 그러나 NeRF는 높은 계산 복잡도와 느린 렌더링 속도로 인해 실시간 응용에 제약이 있다. 이러한 배경에서 3D Gaussian Splatting(3DGS)이 주목받고 있다[5]. 3DGS는 연속적 밀도 표현을 사용하는 NeRF와 달리 이산적인 가우시안 입자 기반 표현을 통해 실시간 렌더링과 고품질 시각화의 균형을 달성하며 차세대 3D 재구성 기술로 자리잡고 있다.

본 기고문은 3DGS 및 관련 최신 연구를 중심으로 기

술의 구조적 특징과 성능 향상 전략을 체계적으로 분석한다. 이를 위해 다음 네 가지 주요 기술 동향을 중심으로 세션을 구성하였다: (1) 단일 이미지 기반 3D 재구성 기술(e.g., AGG, SuGaR), (2) 동적 장면 및 시간적 변화 대응 기술(e.g., 4D-GS), (3) 고속-고정밀 실시간 재구성 기법(e.g., 3DLS), (4) 표현 학습 및 효율적 모델 학습 전략(e.g., Group Training)[6, 7, 8, 9, 10]. 각 세션에서는 해당 주제에 부합하는 대표 모델의 기술적 특징과 성과를 심층적으로 다루며, 3DGS의 현재 한계와 향후 과제를 고찰함으로써, 본 기술의 미래 발전 방향을 제시하고자 한다.

II. A Survey of 3DGS

3DGS는 신경 렌더링 기술의 대표적 진보 중 하나로, NeRF의 후속 기술로서 주목받고 있다. 3DGS는 실시간 렌더링 성능과 고품질 시각화를 동시에 달성하기 위해 가우시안 기반 입자 표현을 도입한 것이 특징이다. NeRF가 연속적인 밀도 표현을 통한 계산 집약적 볼륨 렌더링에 의존하는 반면, 3DGS는 위치, 밀도, 색상, 방향성을 정의하는 이산적인 가우시안 스피릿(splat)을 활용한다. 이 스피릿들은 렌더링 시점에서 이미지 평면에 투영되어 픽셀 색상을 계산하며, 반복적인 광선 샘플링을 제거함으로써 NeRF 대비 수십 배 빠른 렌더링 속도를 제공한다. 이러한 구조는 GPU 기반 병렬 연산에 최적화되어 있다.

3DGS의 핵심 원리는 3D 공간에 분포된 가우시안 스피릿을 이용해 장면을 모델링하는 것이다. 각 스피릿은 공간 행렬로 정의되어 공간적 범위와 방향성을 표현하며, 복잡한 형상을 유연하게 재구성할 수 있다. 렌더링 과정에서는 카메라에서 발사된 광선과 스피릿의 교차를 기반으로 색상과 투명도를 조합하여 최종 이미지를 생성한다. 이 방식은 NeRF의 광선별 반복 샘플링을 대체하며, 학습 과정에서는 가우시안의 위치, 크기, 회전, 투명도, 색상과 같은 파라미터를 E2E(End-to-End) 방식으로 최적화

한다. 이를 통해 렌더링 이미지와 목표 이미지(Ground Truth, GT) 간의 오차를 최소화하며 높은 정확도와 시각적 품질을 보장한다.

NeRF와 비교하였을 때, 3DGS는 여러 장점을 제공한다. 이산적 표현 방식은 메모리 사용량과 계산 부담을 줄여 AR/VR, 인터랙티브 미디어와 같은 실시간 응용에 적합하다. 또한, 복잡한 조명이나 세부 묘사에서 NeRF에 필적하거나 우수한 시각적 품질을 달성한다. 그러나 3DGS는 한계도 존재한다. 현재는 다수의 이미지와 정밀한 카메라 파라미터가 제공되는 경우에만 최적의 성능을 발휘하며, 단일 이미지, 동적 장면, 또는 불완전한 데이터 환경에서는 적용이 어렵다.

이를 보완하기 위해 최근 다양한 후속 연구가 등장하고 있다. 이들은 단일 이미지 기반 3D 재구성, 시간적 변화를 포함한 동적 장면 처리, 희소 데이터에 대한 확장성 강화 등을 목표로 3DGS의 입력 데이터의 한계를 완화하고 표현력을 향상시키는 방향으로 발전하고 있다. 또한, 학습 전략과 표현 최적화 기법의 혁신을 통해 3DGS의 효율성과 응용 가능성을 더욱 확장하고 있으며, 방송, 게임, 몰입형 미디어 등 다양한 분야로의 적용이 기대된다.

본 절에서는 3DGS의 기술적 배경, 구조적 특징 그리고 이론적·실용적 의의를 분석하였다. 이를 바탕으로 후속 절에서는 3DGS를 기반으로 한 주요 기술 동향을 심층 탐구하며, 현재 한계를 극복하고 기술의 적용 범위를 넓히는 최신 연구를 다룬다.

III. Analyzing Main Technologies by Session

초기 3DGS 모델이 보여준 높은 효율성과 뛰어난 시각적 품질을 기반으로, 해당 기술은 다양한 방향으로 빠르게 발전하고 있다. 본 장에서는 3DGS의 최신 기술 동향을 네 가지 주요 세션으로 나누어 분석한다: (1) 단일 이미지 기반 3D 재구성, (2) 동적 장면 및 시간축 재구성, (3)

고속·고정밀 렌더링 기술, (4) 효율적인 학습 구조 및 전략. 각 세션에서는 해당 주제에 속하는 대표 모델들의 구조적 특징, 기술적 기여, 그리고 응용 가능성을 구체적으로 살펴본다.

1. 단일 이미지 기반 3D 재구성

단일 이미지 기반 3D 재구성 기술은 제한된 입력으로부터 고차원의 공간 정보를 유추해야 하므로 본질적으로 잘 정의되지 않은 문제(ill-posed problem)에 해당한다. 그러나 최근에는 사전 학습된 대규모 이미지-3D 매핑 네트워크와 확률적 표현 방식의 결합을 통해, 3DGS 기법을 단일 뷰로 확장하려는 시도가 활발히 이루어지고 있다. 본 절에서는 대표적인 모델인 AGG와 SuGaR를 중심으로 기술적 메커니즘과 기여점을 분석한다[8].

1) AGG: Amortized Gaussian 3D Generation from a Single Image

AGG는 단일 이미지로부터 3D 가우시안을 생성하는 feed-forward 방식의 생성 모델로, NVIDIA와 텍사스 오스틴 대학교의 공동 연구팀에 의해 제안되었다[6]. 기존 3DGS 기반 모델들이 개별 객체에 대해 최적화를 수행하는 방식에 비해, AGG는 사전 학습된 신경망을 통해 단일 이미지 입력으로 3D 가우시안을 즉시 생성한다. 이를 통해 추론 시 최적화 과정 없이도 다양한 객체 카테고리 에 대한 일반화된 3D 재구성을 가능하게 한다. AGG는 두 단계로 구성된 cascaded 생성 파이프라인을 채택하여, 초기에는 저해상도 가우시안을 생성하고, 이후 고해상도 가우시안으로 업샘플링하여 정밀한 3D 모델을 최종적으로 출력한다.

AGG의 핵심은 하이브리드 표현(hybrid representation)을 사용한 두 단계 생성 과정이다.

① Coarse Hybrid Generator

첫 번째 단계에서는 단일 입력 이미지를 DINOv2 비전

트랜스포머를 통해 인코딩하여 특징을 추출한다. 이후, 두 개의 별도 트랜스포머 네트워크가 기하학적 정보와 텍스처 정보를 각각 예측한다:

- Geometry Predictor: 입력 이미지의 특징에서 3D 가우시안의 중심 위치($\mu \in \mathbb{R}^3$)를 예측한다. 트랜스포머는 학습 가능한 쿼리 토큰을 입력으로 받아, Cross Attention과 Self Attention을 통해 위치 시퀀스를 생성한다.
- Texture Field Generator: Triplane을 활용하여 텍스처 필드를 생성하며, Geometry Predictor에서 생성된 3D 위치를 쿼리로 사용하여 색상($c \in \mathbb{R}^3$)과 투명도($\alpha \in \mathbb{R}^1$)를 예측한다. Triplane을 3D 위치에 대한 보간된 특징을 MLP로 디코딩하여 가우시안 속성을 완성한다.

각 가우시안은 위치 μ , 공분산 행렬 Σ , 색상 c , 투명도 α 로 정의되며, 공분산 행렬은 스케일 행렬 $S \in \mathbb{R}^3$ 과 회전 행렬 $R \in \mathbb{R}^3$ 로 분해된다: $\Sigma = RSS^T R^T$. 이때 가우시안 함수는 다음과 같이 표현된다:

$$G(x) = e^{-\frac{1}{2}xy^T\Sigma^{-1}x}$$

최종 픽셀 색상은 α -블렌딩을 통해 계산된다:

$$C = \sum_{i \in N} c_i \alpha_i G(x_i) \prod_{j=1}^{i-1} (1 - \alpha_j G(x_j))$$

학습에는 다중 뷰 재구성을 기반으로 한 손실 함수가 사용되며, LPIPS 및 $L1$ 손실 외에 초기에는 가우시안 의사 레이블을 활용한 Chamfer distance 손실로 warm-up 과정을 수행한다:

$$L_{rendering} = L_{rgb} + w_1 L_{LPIPS} + w_2 L_{Chamfer}$$

② Gaussian Super-Resolution

두 번째 단계에서는 coarse Gaussian을 고해상도로 확장한다. Point-voxel 컨볼루션을 활용한 U-Net 아키텍처

가 사용되며, 입력 이미지의 RGB 정보가 Cross Attention layer를 통해 네트워크에 주입되어 텍스처의 디테일을 정밀화한다. Feature expansion을 통해 가우시안 수를 증가시키며, 최종적으로 고품질 3D 가우시안을 생성한다.

③ Overcoming Challenge in Amortized Training

기존 3DGS에서는 가우시안의 동적 생성 및 삭제(adaptive density control)가 가능하지만, AGG는 feed-forward 네트워크의 특성상 고정된 수의 가우시안을 생성한다. 대신 학습 안정성을 확보하기 위해 등방성 스케일과 회전을 사전에 정의한다. 이후, 초기화 문제를 해결하기 위해 다중 뷰 데이터에서 생성한 가우시안 의사 레이블을 활용하여 Chamfer distance 기반 사전 학습을 수행한다.

AGG는 최적화 기반 3DGS 모델인 DreamGaussian과 비교하였다[11]. DreamGaussian은 Score Distillation Sampling(SDS)을 사용하여 개별 객체에 대한 가우시안을 최적화하지만, AGG는 Amortized 방식으로 단일 이미지에서 즉시 가우시안을 생성한다. AGG는 DreamGaussian 대비 CLIP Distance에서 약간 성능이 떨어지지만(0.3458 vs 0.4293), 300배 이상 빠른 추론 속도(0.19s vs 60s)를 달성하며, 실시간 응용에 훨씬 적합한 구조를 보여준다.

2) SuGaR: Surface Gaussian Reconstruction from a Single Image

SuGaR는 단일 RGB 이미지로부터 정확한 3차원 표현을 복원하고, 이를 3DGS 표현으로 변환하는 hybrid 방식의 3D 재구성 모델이다. 기존의 NeRF 및 3DGS 기반 접근법들이 포인트 클라우드 혹은 밀도 기반 표현을 중심으로 한 반면, SuGaR는 중간 단계로 3D 메시 표현을 직접 복원하고 이를 확장 가능하고 해석 가능한 가우시안 표현으로 변환함으로써, 정밀한 기하 구조와 시각 품질을 동시에 확보한다. SuGaR의 주요 기여는 다음과 같다:

- 표면 정렬을 통해 3D 구조 정확도 향상
- 단일 GPU에서 빠른 메시 생성
- 메시와 가우시안의 공동 최적화로 렌더링 품질 및 편집 효율 개선

SuGaR는 세 단계로 구성된다.

① Aligning the Gaussian with the Surface

가우시안이 장면 표면에 정렬되도록 정규화 항목 R 을 적용한다. 밀도 함수 $d(\mathbf{p})$ 를 정의하고, 가장 가까운 가우시안 \mathbf{g}^* 의 기여도를 기반으로 이상적인 Signed Distance Function(SDF)을 계산한다:

$$R = \frac{1}{|\mathcal{P}|} \sum_{\mathbf{p} \in \mathcal{P}} |\hat{f}(\mathbf{p}) - f(\mathbf{p})|$$

여기서 $\hat{f}(\mathbf{p})$ 는 깊이맵을 활용해 효율적으로 계산된다. 15,000회 반복 학습 중 초기 7,000회는 자유 최적화를 진행하며, 이후 정규화 항목을 적용하여 가우시안을 점진적으로 표면에 정렬한다.

② Efficient Mesh Extraction

학습된 밀도 함수의 특정 레벨셋($\lambda = 0.03$)에서 포인트를 추출하고, 이를 기반으로 Poisson 재구성 알고리즘을 이용해 메시를 생성한다. 이 방식은 흔히 사용되는 Marching Cubes보다 더 정확한 표면 복원이 가능하며, 약 5~10분 내에 메시를 추출할 수 있다.

③ Binding New 3D Gaussian to the Mesh

초기 메시의 각 삼각형에 얇은 가우시안을 바인딩 후, 2D 스케일 조정과 복소수 기반의 회전 최적화를 통해 가우시안의 형태를 정제한다. 이 과정은 2,000회~15,000회 반복되며, 메시의 정합성과 렌더링 결과를 동시에 향상시킨다. 최종적으로 고품질의 렌더링 결과와 편집 가능한 3D 표현을 출력한다.

SuGaR는 3DGS 기반 모델들과 비교하였을 때, 기존 3DGS가 다중 이미지와 긴 최적화 시간을 요구하는 데 비

해 단일 이미지 입력과 단일 GPU 환경에서도 수 분 내 처리가 가능하였다. 또한, 렌더링 품질 측면에서도 PSNR, SSIM 등의 주요 지표에서 3DGS 대비 유사하거나 우수한 성능을 달성하였다. 이러한 성능을 통해 단일 이미지 입력으로도 높은 품질의 3D 메시를 복원하고, 실시간 렌더링에 적합한 3DGS 표현으로 전환할 수 있는 모델임을 확인하였다.

2. 동적 장면 및 시간축 재구성

정적 장면을 대상으로 기존 3DGS 기반 모델은 시간축 변화나 동적 객체의 처리에 제약이 있으며, 이는 실시간 미디어와 같은 응용 분야에서 한계로 작용한다. 이러한 제약을 극복하기 위해 최근에는 시간축을 포함하는 4D 표현(Four-Dimensional Scene Representation)을 활용한 연구들이 활발히 진행되고 있다. 본 절에서는 대표적으로 4D Gaussian Splatting(4D-GS) 모델을 중심으로, 동적 장면 재구성 기술의 구조와 성능을 분석한다.

1) 4D-GS: 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering

4D-GS는 시간적으로 변화하는 장면을 고효율적으로 표현하고 실시간 렌더링을 가능하게 하는 장면 표현 방식이다. 기존 방식은 각 시간 프레임마다 3D 가우시안을 독립적으로 구축하거나, 시간축을 따라 동일한 정보가 중복되어 메모리 사용량이 급증하는 단점이 있다. 반면 4D-GS는 시간에 따라 움직이고 변형되는 장면을 단일한 정준(canonical) 3D 가우시안 세트와 가우시안 변형 필드 네트워크로 표현하여 계산 및 저장 효율을 향상시킨다.

4D-GS에서 사용하는 Canonical 3D Gaussian은 각 가우시안마다 위치, 크기, 방향, 색상 및 불투명도 정보를 포함한다. 초기에는 Structure-from-Motion(SfM)을 이용하여 정적 장면 기반으로 3D 가우시안을 생성한 후, 학습을 통해 움직임이 있는 부분에 집중하도록 최적화한

다. 이후, Gaussian Deformation Field Network를 통해 Canonical 가우시안을 시간 t 에 따른 변형된 가우시안 G' 으로 매핑한다. 이 네트워크는 시간 및 공간 정보를 동시에 인코딩하는 1) Spatial-Temporal Structure Encoder $H(G, t)$ 와, 해당 정보를 이용해 위치, 회전, 크기 등의 변형을 예측하는 2) Multi-head Gaussian Deformation Decoder로 구성된다. 변형된 3D 가우시안 G' 을 이용하여 이미지 평면상에 직접 투영하여 렌더링 이미지를 생성한다.

① Spatial-Temporal Structure Encoder

가우시안의 위치 및 시간 정보를 활용해 6개의 2차원 voxel plane으로부터 특징을 추출한다. 공간 차원은 (x, y) , (x, z) , (y, z) 로 나타내며, 시간 차원은 (x, t) , (y, t) , (z, t) 로 나타낸다. 각 voxel plane은 멀티 해상도로 구성되며, voxel query 결과는 작은 MLP를 통해 통합된다. 이를 통해 주변 가우시안 간의 공간적, 시간적 연관성을 반영한 고차원 특징 벡터를 생성한다.

② Multi-head Gaussian Deformation Decoder

인코더에서 추출된 특징 벡터를 기반으로, Multi-head 구조의 MLP 디코더를 통해 다음과 같은 세 가지 요소의 변형값을 예측한다: 위치 변형 ($\Delta x, \Delta y, \Delta z$), 회전 변형 (Δr), 스케일 변형 (Δs). 이를 통해 시간 t 에 해당하는 최종 3D 가우시안 $G' = G + \Delta G$ 을 얻고, 이를 렌더링에 사용한다.

기존 3DGS는 시간 변화가 없는 정적 가우시안 표현만 사용하기 때문에, 동적 장면에 적용할 경우 각 시간 프레임마다 별도의 모델을 구성해야 하며, 이로 인해 메모리 사용량이 커지고 재사용성이 떨어진다. 반면 4D-GS는 Canonical 가우시안과 시간에 따른 변형을 통해 동일한 파라미터 공간 내에서 장면의 동적 특성을 효과적으로 모델링할 수 있다. 이로 인해 메모리 효율성 증가와 학습 효율성 및 렌더링 품질 성능이 향상되는 장점

이 나타난다.

3. 고속·고정밀 렌더링 기술

실시간 3D 콘텐츠 제작에서는 고품질 이미지를 빠른 속도로 생성할 수 있는 렌더링 기술이 필수적이다. 기존 3DGS는 사실적인 시각 품질을 제공하지만, 고속 렌더링 측면에서는 계산량과 메모리 사용량이 큰 제약으로 작용한다. 본 절에서는 이러한 문제를 해결하고자 제안된 3DLS 모델을 중심으로, 고속성과 고정밀성을 추구하는 최신 기술을 소개한다.

1) 3DLS: 3D Gaussian Splatting with Learnable Shape Embedding

3D Linear Splatting(3DLS)은 기존 3DGS 구조를 개선하여, 선형 커널(linear kernel)을 도입함으로써 고주파 영역의 세부 사항을 더 정밀하게 포착할 수 있도록 설계된 렌더링 기법이다. 기존 3DGS는 각 3D 포인트를 위치 X , 공분산 Σ , 색상, 투명도를 포함하는 가우시안으로 표현하고, 이를 2D 이미지로 투영하여 렌더링한다. 이때 가우시안 커널의 부드러운 경계는 사실적인 효과를 제공하지만, 고주파 텍스처나 뚜렷한 경계 표현에서는 한계가 있다.

이에 반해 3DLS는 지수 함수 대신 선형 보간 기반의 커널을 도입하여, 더 선명한 경계와 정확한 텍스처 재현이 가능하다.

3DLS는 기존 3DGS의 품질적·효율적 한계를 해결하기 위해 다음과 같은 세 가지 측면에서 기술적 발전을 도모하였다:

① Transition to Linear Kernels

기존 3DGS는 공분산 행렬 $\Sigma = RSS^T R^T$ 에 기반한 가우시안 함수를 사용하여, 픽셀 기여도를 계산하는 방식이다. 하지만 이 방식은 경계선, 세부 텍스처와 같은 고주파 성분을 뚜렷하게 표현하기 어렵다. 3DLS는 지수 함수

대신 선형 보간 커널을 도입하여, 이미지 상에서 픽셀 기여도를 선형적으로 계산한다. 이를 통해 경계가 더 뚜렷해지고, floating artifacts가 줄어들며, 결과적으로 고해상도 텍스처가 더 정확히 재현된다.

② Training Optimization

3DLS는 $L1$ 색상 손실과 LPIPS 손실을 조합한 손실 함수를 사용하여 시각적으로 더욱 세밀한 학습이 가능하도록 했다. 또한, 초기 포인트 배치는 SfM 기반 포인트를 활용해 안정적인 학습을 보장하며, 밀도 제어 전략을 통해 불필요한 가우시안 입자를 제거하여 효율성을 높였다.

3DLS는 고주파 세부 표현, 실시간 속도, 효율적인 저장 구조 측면에서 기존 3DGS를 뛰어넘는 성능을 보여준다. 특히, 고해상도 텍스처와 복잡한 기하 구조를 다루는 장면에서 우수한 재구성능을 제공한다. 이는 AR/VR 콘텐츠 제작 등에서 실시간 고품질 3D 렌더링을 위한 유력한 기술 대안이 된다. 다만, 현재 버전의 3DLS는 정적 장면에서 최적화되어 있어서 동적 객체 표현에는 한계가 있다. 이러한 점을 해결하기 위해서는 4D-GS 등과의 통합적 연구가 향후 과제로 제시된다.

4. 효율적인 모델 학습 전략

3DGS는 Novel View Synthesis(NVS) 분야에서 고품질 실시간 렌더링을 가능하게 하여 다양한 응용 분야에서 활용되고 있다. 그러나 수백만 개의 가우시안 프리미티브가 요구하는 높은 계산량과 학습 시간은 실용적 제약으로 작용한다. 이를 해결하기 위해 최근 제안된 Group Training 기법은 학습 속도를 최대 30% 향상시키는 동시에, 렌더링 품질을 유지 또는 개선할 수 있는 전략이다. 본 절에서는 Group Training의 구조 및 방법론을 소개하고, 3DGS 기반의 기존 학습 방법과 비교하여 그 효율성과 성능 개선 효과를 기술한다.

1) Faster and Better 3D Splatting via Group Training

Group Training은 전체 가우시안 프리미티브를 학습 중 동적으로 두 개의 그룹으로 나누어 관리한다. Under-training Group(UTG)은 현재 학습에 적극적으로 참여하는 가우시안 집합으로, 렌더링 및 최적화에 직접 기여한다. Caching Group(CG)은 학습 중 일시적으로 제외된 프리미티브로, 메모리와 계산 자원을 절약하는 데 기여한다. 이러한 그룹 분리는 프리미티브의 수가 기하급수적으로 증가하는 기존의 densification 과정에서 발생하는 계산 병목을 완화하며, 전체 모델의 효율성과 확장성을 높인다. 또한, Group Training은 기존의 3DGS 또는 Mip-Splatting 구조와도 높은 호환성을 지닌다.

Group Training의 핵심은 가우시안 프리미티브를 효과적으로 샘플링하고, 주기적으로 그룹을 재구성하는 방식이다. 두 가지 주요 샘플링 전략이 사용된다:

- Random Sampling(RS): 단순 무작위 샘플링을 통해 UTG를 구성하는 방식으로 구현이 간단하나, 중요도가 낮은 프리미티브가 포함될 수 있어 효율성 면에서 제한적이다.
- Opacity-based Prioritized Sampling(OPS): 프리미티브의 투명도 α_i 를 기준으로 중요도를 부여하고, 높은 투명도의 프리미티브를 우선적으로 학습에 포함시킨다. OPS의 샘플링 확률은 다음과 같이 정의된다:

$$P_i = \frac{\alpha_i}{\sum_{i=1}^N \alpha_i}$$

OPS는 α -블렌딩 과정에서 빠른 투명도 포화(transmittance saturation)를 유도하여 렌더링 속도 향상과 시각 품질 유지를 동시에 달성할 수 있는 근거를 갖는다.

Group Training은 기본적으로 500회 반복마다 그룹을 재구성하며, 전체 학습 주기는 다음과 같은 단계

로 이루어진다. 먼저, UTG와 CG를 병합하여 전체 가우시안을 수집하고, 병합된 가우시안을 이용해 학습 뷰를 렌더링한다. 이후, OPS 또는 RS를 적용하여 새로운 UTG와 CG를 생성하며, 마지막으로 UTG만을 대상으로 densification 및 최적화를 수행한다. 이러한 주기적 갱신은 CG를 통해 계산 부담을 줄이면서도, 전체 가우시안이 학습에 골고루 참여할 수 있도록 하여 학습 품질 저하를 방지한다. 투명도 중심의 프리미티브 중요도 기반 최적화는 복잡한 장면에서도 컴팩트한 모델 생성을 가능케 하여, 향후 대규모 데이터셋이나 모바일 환경에서의 3DGS 응용에 강점을 가진다. 다만, 극단적으로 동적인 장면이나 초대형 시나리오에서는 추가적인 확장 연구가 필요하다.

IV. Challenges and Future directions

3DGS는 고품질의 실시간 렌더링을 가능하게 하며 NVS 분야에서 빠르게 발전해 왔지만, 여전히 해결되지 않은 기술적 과제가 존재한다. 첫째, 복잡한 동적 장면에서의 표현력은 제한적이다. 4D-GS와 같은 확장은 시간축을 도입했음에도 비선형 동작이나 비강체 변형을 완전하게 처리하지 못하며, 실시간 응용에는 제약이 있다. 둘째, 수백만 개의 가우시안 프리미티브를 다루는 과정에서 발생하는 메모리 및 연산 자원 부담은 모바일 환경이나 대규모 장면 적용에 장애물로 작용한다. 또한, 다양한 장면 조건에서의 일반화 성능이 부족하고, 현재 대부분의 프레임워크는 장면의 의미적 정보를 반영하지 않아 상호작용 기반 응용에는 한계가 있다. 이러한 문제를 해결하기 위해, 향후 연구는 비선형 시간 모델링, 다양한 도메인에 강인한 일반화 기법, 그리고 의미 기반 3D 표현 통합 등으로 나아가야 한다. 아울러, 데이터 수집부터 렌더링까지를 포함하는 E2E 실시간 파이프라인 구축 연구 역시 실용화를 위한 핵심 과제로 제시된다. 이와 같은 방향은 3DGS의 활용 영

역을 방송 미디어, 메타버스, 자율 주행 등으로 확대시키는 중요한 토대가 될 것이다.

V. 결론

3DGS는 NeRF의 한계를 넘어 실시간 3D 장면 재구성을 가능하게 하는 차세대 기술이다. 이산적인 가우시안 입자를 활용해 빠른 처리와 고품질 시각화를 동시에 달성하여, 메타버스 및 증강현실 등에서 큰 가능성을 보여준다. AGG와 SuGaR는 단일 이미지로 정밀한 3D 모델을 만들고, 4D-GS는 시간축을 포함한 동적 장면을 효율적으로

렌더링한다. 또한, 3DLS는 고주파 텍스처를 섬세하게 표현하면서도 속도를 높인다. 마지막으로 Group Training 기법은 학습 효율성을 개선해 대규모 모델의 실용성을 강화한다. 하지만, 3DGS에도 아직 해결해야 할 과제가 남아 있다. 특히, 비선형 움직임이나 복잡한 동적 장면을 표현하는 데 한계가 있으며, 모바일 환경에서 자원을 효율적으로 관리하는 문제가 남아 있다. 앞으로의 연구는 비강체 움직임 모델링과 다양한 도메인에 적용 가능한 학습 전략, 시맨틱 정보를 활용한 표현 학습으로 확장될 필요가 있다. 이러한 발전은 3DGS가 단순한 연구를 넘어 실제 산업 현장에서 활용될 수 있는 토대를 마련하고, 실감형 콘텐츠 제작의 혁신을 이끌어낼 것으로 기대된다.

참 고 문 헌

- [1] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* 1 (2006), pp. 519-528.
- [2] J.L. Schonberger and J.M. Frahm, "Structure-from-Motion Revisited," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 4104-4113.
- [3] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics & Automation Magazine* 13 (2006), pp. 99-110.
- [4] 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," *Communications of the ACM* 65 (2021), pp. 99-106.
- [5] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics* 42 (2023), Article No. 139.
- [6] D. Xu, Y. Yuan, M. Mardani, S. Liu, J. Song, Z. Wang, and A. Vahdat, "AGG: Amortized Generative 3D Gaussians for Single Image to 3D," *arXiv preprint arXiv:2401.04099* (2024).
- [7] A. Guédon and V. Lepetit, "SUGAR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 5354-5363.
- [8] G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, et al., "4D Gaussian Splatting for Real-Time Dynamic Scene Rendering," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 20310-20320.
- [9] H. Chen, R. Chen, Q. Qu, Z. Wang, T. Liu, X. Chen, and Y.Y. Chung, "Beyond Gaussians: Fast and High-Fidelity 3D Splatting with Linear Kernels," *arXiv preprint arXiv:2411.12440* (2024).
- [10] C. Wang, G. Ma, Y. Xue, and Y. Lao, "Faster and Better 3D Splatting via Group Training," *arXiv preprint arXiv:2412.07608* (2024).
- [11] J. Tang, J. Ren, H. Zhou, Z. Liu, and G. Zeng, "DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation," *arXiv preprint arXiv:2309.16653* (2023).

저 자 소 개



박성준

- 2025년 2월 : 동아대학교 AI학과(공학사)
- 2025년 ~ 현재 : 동아대학교 컴퓨터공학과 석사과정
- 주관심분야 : 멀티미디어, 딥러닝, 머신 비전, 3차원 비전, 생성형 AI



조수빈

- 2022년 ~ 현재 : 동아대학교 컴퓨터공학과 학사과정
- 주관심분야 : 멀티미디어, 딥러닝, 머신 비전, 3차원 비전, 생성형 AI



서정일

- 2005년 2월 : 경북대학교 전자공학과(공학박사)
- 1998년 ~ 2000년 : LG반도체 선임연구원
- 2000년 ~ 2023년 : 한국전자통신연구원 실감미디어연구실장
- 2023년 ~ 현재 : 동아대학교 컴퓨터공학과 부교수
- 주관심분야 : 멀티미디어, 오디오/비디오 부호화, 딥러닝, 머신 비전