

Recent Advances in Motion Modeling for Dynamic 3D Gaussian Splatting

2026 통계 세미나



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

최승규

Outline

- Background
 - Novel View Synthesis
 - 3DGS 이후의 연구
 - 논문 선정 동기
- LocalDyGS: Multi-view Global Dynamic Scene Modeling via Adaptive Local Implicit Feature Decoupling [ICCV 2025]
- ReCon-GS: Continuum-Preserved Gaussian Streaming for Fast and Compact Reconstruction of Dynamic Scenes [NeurIPS 2025]

Background

- Novel View Synthesis
 - Scene
 - 표현하고자 하는 장면 전체
 - Frame
 - 시간의 흐름에 따른 장면의 순간적인 상태
 - View
 - 특정 위치와 각도에서 장면을 바라보는 카메라의 시점
 - Monocular vs Multi-View
 - Novel View Synthesis:
 - 학습에 사용되지 않은 시점(View)의 장면을 합성해내는 Task



Background

- Novel View Synthesis

- Point Cloud

- 3D 공간 상의 점들로 장면을 표현
- 확대하거나 특정 각도에서 보면 점 사이가 비어 형태가 불분명

- Mesh

- Point Cloud의 점들을 선으로 연결하여 다각형 면을 형성한 구조

- NeRF¹⁾

- MLP 신경망 사용
- 고화질 출력이 가능하지만 학습이나 렌더링 속도가 매우 느림

- 3D Gaussian Splatting²⁾

- NeRF의 속도 문제를 해결하며 등장한 게임 체인저 (30 FPS+)
- 3D 공간에 가우시안 타원체를 뿌리고 2D로 projection
- NeRF와 같은 ray-marching 방식 대신 CUDA Rasterizer를 통한 투영 방식 채택
- 이미지넷 챌린지의 CNN과 같은 혁신!

Background

- Novel View Synthesis

- 3D Gaussian Splatting¹⁾의 파라미터 구조

- Position (μ)

- ☼ 가우시안 타원체의 중심 위치

- Covariance (Σ)

- ☼ 가우시안 타원체의 크기와 회전

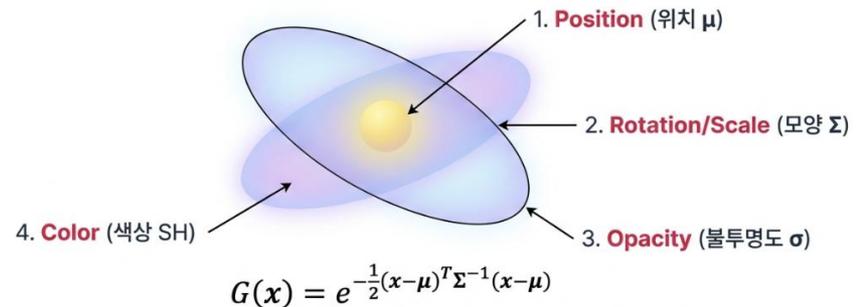
- ☼ 대각 행렬 형식의 크기 행렬 s 와 회전 쿼터니언 q 를 각각 학습 후 결합

- Opacity (α)

- ☼ 불투명도

- Color (Spherical Harmonics)

- ☼ 시점 의존적 색상 표현



Background

- 3DGS 이후의 연구

- In-the-wild

- 다양한 조명 조건 하에서 촬영된 장면의 색감을 통합

- Continual Learning

- 이전 학습 내용을 유지하며 새로 추가된 후속 장면의 정보를 통합

- Compression

- 렌더링 퀄리티를 유지하며 저장 용량을 줄이고 FPS 향상

- Dynamic Scene Reconstruction

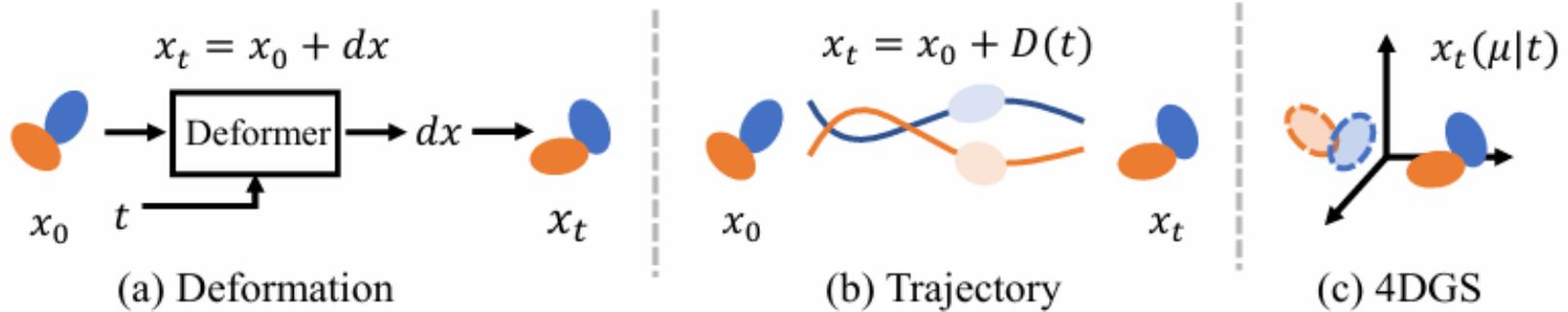
- 정적 장면의 재구성에 그치지 않고 시간의 흐름에 따라 변하는 장면을 재구성
 - 오늘 세미나의 주제!



Background

- 논문 선정 동기

- 단순히 여러 프레임을 따로 학습하면 저장 공간 폭발, 학습 속도 저하
- Motion에 대한 별도 고려가 없으면 각종 Artifact 발생
- 최신 논문들이 이 “Motion Representation”을 어떻게 해결했는지 유형 별로 분석



**LocalDyGS:
Multi-view Global Dynamic Scene Modeling
via Adaptive Local Implicit Feature Decoupling
[ICCV 2025]**

Introduction

- 기존 연구의 문제점
 - Monocular Dynamic Scene Reconstruction
 - 간단한 장면에만 제한적
 - 크게 동적이고 복잡한 장면은 어려움
 - Implicit MLP
 - 단일 MLP를 통한 전체 동적 장면의 표현 시도 -> 용량 부족
 - 크고 복잡한 모션에서 표현력 부족으로 인한 Blurring, Deblurring
 - Explicit Motion Function
 - 각 가우시안에 다항식 궤적 할당 -> 학습이 매우 느리고 파라미터가 너무 많음
 - 4D Primitives
 - 가우시안 개수가 폭발적으로 증가하여 무거움
- 문제 해결 방법
 - Global Space를 Local Space로 분해
 - 각 Local Space 내부의 움직임 표현을 위해 Temporal Gaussians 생성 (MLP 이용)

Method

- Preliminary: 3DGS & Scaffold-GS

- 3DGS

- 위치, 회전, 크기, 불투명도, 색상을 가지는 3D 가우시안을 2D 화면에 투영
 - $G(x) = e^{-\frac{1}{2}(x)^T \Sigma^{-1}(x)}$

- Scaffold-GS

- Anchor를 기준으로 Neural Gaussian을 생성
 - 정적 장면에만 한정

- LocalDyGS의 차별점

- 정적 장면의 재구성에 그치지 않고 Anchor 개념을 시공간으로 확장
 - Seed를 통해 동적 장면을 효율적으로 표현

Method

- Global Seeds Initialization

- Local Space를 커버하는 Seed의 위치 초기화

- 초기 프레임의 SfM 포인트 클라우드를 병합, 초기 Seed 위치로 사용
 - 물체가 나타날 수 있는 곳에 대한 사전 정보 제공 역할

- Seed 파라미터

- Seed는 시간이 흘러도 변하지 않는 정적 정보를 포착, 64차원 벡터로 저장
 - Local Space의 커버 범위를 나타내는 3차원 스케일 파라미터 v

- 스케일 파라미터의 초기화

- 가장 가까운 3개 Seed까지의 거리를 평균
 - Seed 간 거리가 넓어지면 커버 범위가 넓어짐

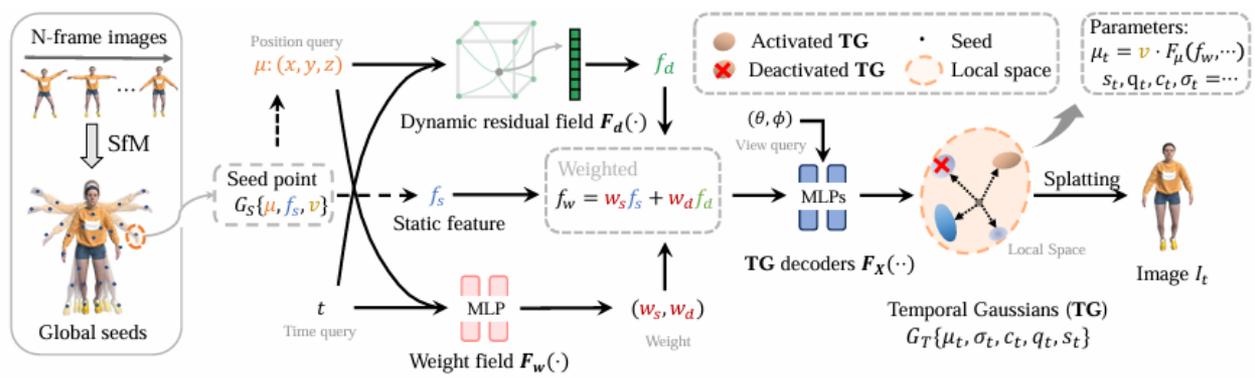
Methods

- Feature-Decoupled Spatio-Temporal Fields

- 정적 정보와 동적 정보를 분리하여 효율적으로 모델링
- 단일 MLP로 시공간을 한 번에 배우려고 하면 용량 한계로 artifact 발생
 - 정적인 특징과 동적인 특징을 분리하여 학습 후 병합
- Dynamic Residual Field (F_d)
 - 입력: 시간, 위치
 - 출력: 동적 잔차 특징 f_d
- Adaptive Weight Field (F_w)
 - 입력: f_s, f_d
 - 출력: 두 모델의 가중치
- 최종 특징: $w_s * f_s + w_d * f_d$

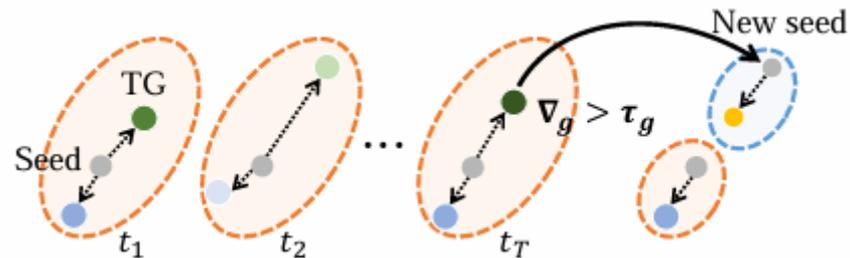
Methods

- Local Temporal Gaussian Derivation
 - 결합된 특징을 바탕으로 실제 렌더링 가능한 Temporal Gaussians 생성
 - 2-layer MLP가 Gaussian 특성 별로 존재
 - 하이퍼파라미터 K를 통해 시드 당 출력 가우시안 수 조정
 - q, s, alpha, c
 - Deactivation Strategy
 - 출력된 가우시안 중 불투명도가 임계값 이하인 가우시안은 렌더링 시 제외



Methods

- Adaptive Seed Growing (ASG)
 - 초기 포인트 클라우드의 불완전함을 학습 중 보완
 - 초기 SfM 포인트가 없는 곳은 시드가 존재하지 않고 텅 비어있음
 - 기울기가 임계값보다 큰 가우시안은 위치가 불안정하다고 판단
 - 해당 위치에 새로운 Seed를 추가하고 Local Space를 생성
 - 빈 공간이 점진적으로 채워지며 렌더링 품질 향상



Methods

- Loss Function

- 3DGS의 Loss 그대로 차용

- L1 loss: 픽셀 간 차이를 뺄셈
 - L_SSIM: 시각적 유사도를 고려

- Volume Regularization

- 가우시안이 불필요하게 커져 다른 Local Space를 침범하는 것을 방지
 - 스케일의 곱을 통해 계산

- 최종 Loss

- $L = (1 - \lambda) * L_1 + \lambda * L_{SSIM} + \lambda_v * L_v$

Experiments

- Quantitative Results

- N3DV

- PSNR 32.28dB
 - SpaceTimeGS 대비 학습 시간 1/10 (35분), 저장 공간 1/2 (100MB)
 - 농구공, 커피 줄기 등 포착하기 어려운 motion도 효율적 포착

- Ablation Study

- w/o Static Feature

- 배경이 흔들리고 화질이 저하됨

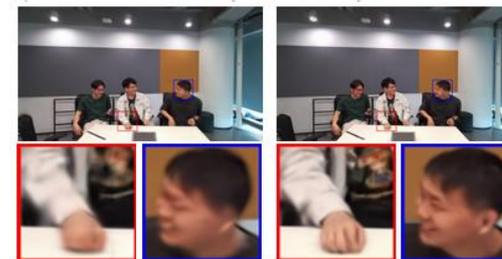
- w/o ASG

- 텍스처가 없는 지형 표현 불가



(a) GT

(b) Ours



(a) w/o ASG

(b) w/ ASG

**ReCon-GS:
Continuum-Preserved Gaussian Streaming for
Fast and Compact Reconstruction of Dynamic Scenes
[NeurIPS 2025]**

Introduction

- 기존 연구의 문제점
 - Online / Streaming
 - 실시간 처리는 가능하지만 오차가 누적됨 (Drift)
 - Deformation Field
 - 물리적인 구조를 무시
 - 파라미터 낭비, 정확도 저하
- 해결 방법
 - Motion을 Coarse-to-fine으로 3단계 분해
 - 주기적으로 Anchor를 재설정하여 Drift 방지
 - Anchor가 바뀌어도 기존의 기하 정보를 수학적으로 상속

Method

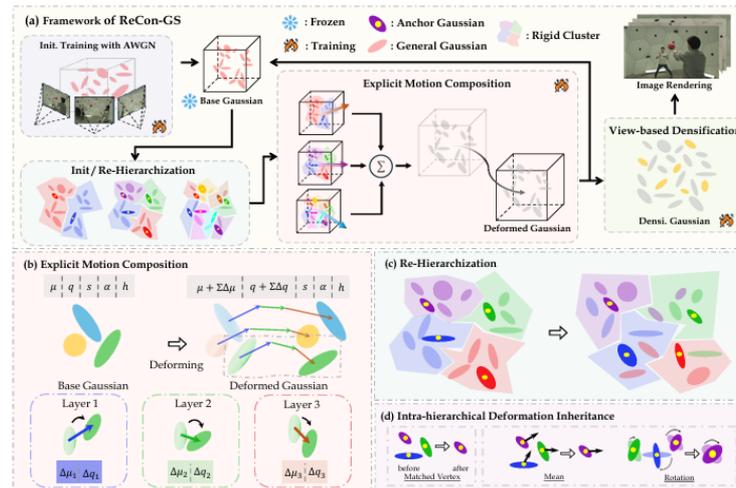
- Overview

- Base Gaussians

- 첫 번째 프레임을 통해 정적 3DGS 장면 학습
- 위치 정보에 노이즈를 섞어 가우시안들이 더 고르게 분포하게 유도

- Streaming Process

- 이후 비디오 프레임에 대해서는 기하학적 형태만 변형시킴
- 색상, 불투명도 등의 외관 특징은 계속 고정



Method

- Adaptively Hierarchical Motion Representation

- Anchor Gaussian Initialization

$$M = \lceil N_{anchor}^{1/3} \rceil$$

$$g_a = \arg \min_{g \in \mathcal{C}_{i,j,k}} \|\mu - c_{i,j,k}\|_2, \text{ where } c_{i,j,k} = \mu_{\min} + \left(i + \frac{1}{2}, j + \frac{1}{2}, k + \frac{1}{2}\right) \odot \frac{\Delta}{M}$$

- 공간을 그리드로 나누고, 각 셀 중심에 가장 가까운 가우시안을 앵커로 선정

- Hierarchical Rigid-Cluster Formation

$$M = \lceil (N_{anchor} \cdot 3^{l-1})^{1/3} \rceil$$

$$g_a^* = \arg \min_{g_a \in \mathcal{G}_A} \|\mu_n - \mu_a\|_1$$

- 상위 레벨에서 하위 레벨로 갈 수록 촘촘하게 그리드를 나눔 (Coarse-to-fine)
 - 각 가우시안은 Anchor이거나, 가장 가까운 Anchor에 종속되는 General Gaussian

- Explicit Motion Composition

$$\Delta\mu_h = \sum_{l=1}^3 \Delta\mu^{(l)}, \quad \Delta q_h = \sum_{l=1}^3 \Delta q^{(l)}$$

- MLP 통과 없이 단순 덧셈 연산, 빠른 속도
 - “미소 변화량” 으로서 정의하고 학습하므로 쿼터니언 덧셈에 문제가 없음

Method

- Dynamic Hierarchy Reconfiguration

- Periodic Re-Hierarchization

- Streaming 방식 특유의 Drift 문제를 해결하기 위함
- 주기적으로 현 상태의 가우시안 위치를 반영하여 Anchor 재선정

- Intra-hierarchical Deformation Inheritance

- 새로운 Anchor는 초기화 시 변형 파라미터 값이 0이므로 이전 기하 정보 상속 필요

- 이동 상속

- ※ 단순 좌표 값이므로 단순 합산
$$\Delta\mu'_a^{(l)} = \frac{1}{3} \sum_{i=1}^3 \Delta\mu_{a_i}^{(l)}$$

- 회전 상속

- ※ 쿼터니언은 Double Cover 문제로 단순 평균 불가

- ※ 세 쿼터니언의 “가장 지배적인 방향“ 계산 -> M의 가장 큰 고유값의 고유벡터

- ※ 계산 증명을 위해 레일리 몫, 라그랑주 상수 개념 필요

$$\Delta q'_a^{(l)} = \frac{v_{\max}(\mathcal{M})}{\|v_{\max}(\mathcal{M})\|}, \quad \mathcal{M} = \sum_{i=1}^3 \Delta q_{a_i}^{(l)} (\Delta q_{a_i}^{(l)})^T$$

Method

- Storage-aware Optimization

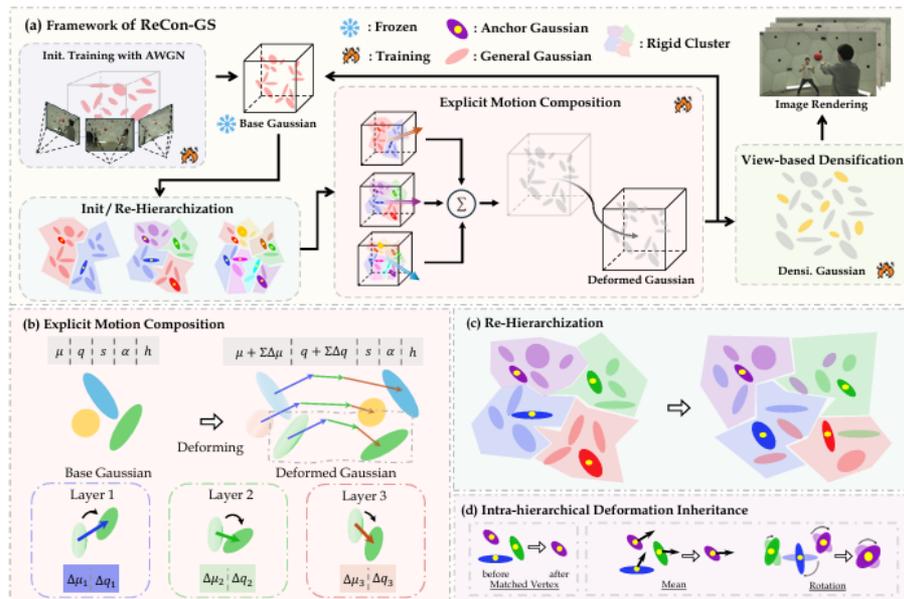
- Density Control

- 사용 가능한 앵커의 수에 따라 밀도 조절 가능

- Two-Phase Optimization

- Phase 1: 앵커의 움직임만 학습, Base Gaussian 고정

- Phase 2: 움직임만으로 표현되지 않는 장면 표현을 위해 Clone/Split/Prune



Experiments

- Quantitative Results

- 기존 Streaming SOTA (HiCOM) 대비 용량 50% 절감 (0.4MB/Frame)
- PSNR 0.5dB 이상 향상
- 250 FPS
- Train Time 6.4s/Frame

- Ablation Study

- Hierarchy Level: 3일 때 최적
- w/o Reconfiguration
 - 시간 경과에 따른 Drift 심각

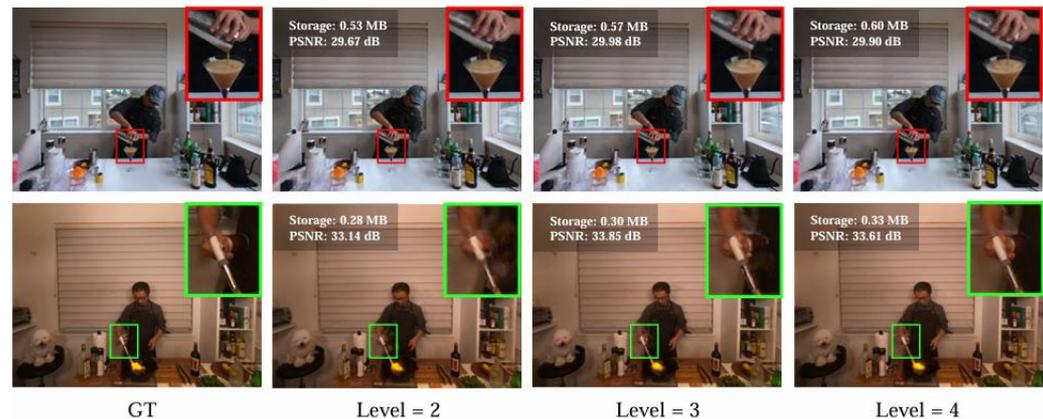


Figure 5: Qualitative results of our ReCon-GS under different hierarchical levels

감사합니다