

## 사용자 인터랙션을 위한 3D Gaussian Splatting 기반 객체 자세 추정 연구

송복득, 김성훈  
한국전자통신연구원

bdsong@etri.re.kr, steve-kim@etri.re.kr

### A Study on Object Pose Estimation Based on 3D Gaussian Splatting for User Interaction

Bok Deuk Song, Sung-Hoon Kim

Hyper-Reality Metaverse Research Laboratory  
Electronics and Telecommunications Research Institute (ETRI).

#### 요 약

본 논문은 3D Gaussian Splatting(3DGS) 기술을 기반으로, 사용자 인터랙션을 접목하기 위해 실시간으로 변화하는 3차원 객체의 자세(pose)를 정밀하게 추정하고 이를 동적으로 반영하는 방법을 제안한다. 제안하는 기법은 정적인 객체로부터 생성된 초기 3DGS 장면을 기반으로 하며, 객체에 이동 또는 회전이 발생했을 때 새로운 시점에서 변화된 객체의 자세를 추정한다. 이를 위해 기존 Gaussian 들과 새롭게 획득된 Gaussian 들 간의 평균 벡터(mean vector) 차이를 계산하여 회전(Rotation) 및 이동(Translation) 파라미터를 도출하며, 각 Gaussian 에 고유 식별자(identifier)를 부여함으로써 정합성과 일관된 대응 관계를 유지한다. 추정된 자세 정보는 기존 3DGS 장면 내 Gaussian 들의 위치와 방향을 동적으로 갱신하는 데 활용되며, 이를 통해 객체의 움직임을 실시간으로 반영할 수 있다. 본 연구는 3D 영상 콘텐츠 제작, 실감형 인터랙티브 미디어, 증강현실(AR) 및 가상현실(VR) 시스템 등 사용자 인터랙션이 요구되는 다양한 응용 분야에서 효과적으로 활용될 수 있을 것으로 기대된다.

#### I. 서 론

3차원 객체의 자세(Pose)를 정밀하게 추정하는 기술은 증강현실(AR), 가상현실(VR), 로봇 비전, 디지털 트윈 등 다양한 실감형 콘텐츠 및 공간 인식 기반 애플리케이션에서 핵심적인 역할을 수행한다. 특히 객체가 움직이거나 회전하는 동적인 환경에서, 해당 객체의 위치 및 방향을 실시간으로 추정하고 이를 반영하는 기술은 고품질 3D 영상 구현과 몰입형 입체미디어 제공에 필수적이다. 최근 들어 3D Gaussian Splatting (3DGS) 기술[1]은 연산 효율성과 렌더링 품질 측면에서 주목받으며, 신속한 3D 장면 재구성과 표현에 적합한 방식으로 평가받고 있다. 그러나 기존의 3DGS 기반 연구들은 대부분 정적인 장면 구성에 초점을 맞추고 있으며, 객체가 시점 변화나 자세 변화를 겪는 상황에서의 동적 업데이트를 다루는 연구는 상대적으로 부족하다. 특히, 움직이는 객체를 실시간으로 반영하려면 기존 Gaussian 들의 위치 및 방향을 빠르게 보정할 수 있는 효율적인 자세 추정 방법이 요구된다.

이에 본 논문에서는 서로 다른 시점에서 획득된 두 장의 컬러 영상을 활용하여 객체의 포즈 변화를 추정하고, Gaussian 들의 평균 위치(mean vector) 차이를 기반으로 회전 및 이동 파라미터를 계산하는 방법을 제안한다.

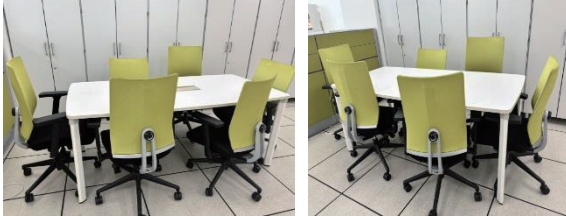
#### II. 본론

본 연구에서는 기존의 방법인 Splatt3R [2]을 기반으로, 두 장의 자연 이미지로부터 사전 카메라 보정이나 깊이 정보 없이 직접 3차원 장면을 재구성하고 새로운 시점을 생성하는 실험을 수행하였다. 아울러, 객체의 자세 변화를 효과적으로 반영하기 위한 3D Gaussian 기반의 자세 추정 방법을 제안한다. 제안하는 방법은 서로 다른 두 시점에서 관측된 3D Gaussian 분포 간의 정합 과정을 통해 회전(Rotation) 및 이동(Translation) 파라미터를 추정하고, 이를 바탕으로 전체 Gaussian 장면의 포즈를 실시간으로 보정하는 방법을 제안한다.

##### 2.1 Splatt3R 을 활용한 3D Gaussian 장면 재구성

Splatt3R 은 사전 카메라 보정 없이 두 장의 자연 이미지(스테레오 이미지)만으로 3차원 장면을 재구성할 수 있는 제로샷 기반의 딥러닝 모델이다. 본 방법은 MAST3R 에서 예측된 3D 포인트 클라우드를 기반으로, 각 픽셀에 대응되는 3D Gaussian primitive 의 위치, 회전, 스케일, 불투명도, 색상 등의 속성을 예측하여 전체 장면을 Gaussian 분포의 집합으로 표현한다. 예측된 Gaussian 들은 실시간 렌더링이 가능하며, 새로운 시점에서도 장면을 자연스럽게 재구성할 수 있도록 한다. 이로써 Splatt3R 은 복잡한 다중 뷰 또는 카메라

파라미터 없이도 효율적인 3D 장면 표현과 시점 생성이 가능하도록 지원한다. 그림 1은 보정 없이 촬영된 이미지 2장을 이용하여 3D gaussian 장면을 재구성한 결과화면이다.



(a) 원본이미지 - 1 (b)원본이미지 - 2



(c)3D Gaussian 장면 재구성 결과 화면

그림 1. Splatt3R을 활용한 3D Gaussian 장면 재구성 결과 화면

## 2.2 Gaussian Mean Vector 추출 및 정렬

본 논문에서 풀고자 하는 문제는 다음과 같은 범함수  $F$ 을 최소화하는 어파인 변환  $Aff(\cdot)$ 를 찾는 최적화 문제로 표현이 된다:

$$\min_{Aff(\cdot)} \mathcal{F} = \sum_v \left| \mathcal{R} \left( \left\{ Aff(G_i^{obj}) \right\} \right) - I^v \right| \quad (1)$$

여기서  $G_i^{obj}$ 는 객체의 표현에 사용되는 Gaussian 들이고,  $\mathcal{R}(\cdot)$ 는 3D Gaussian 을 2D 평면으로 투영하는 함수이고,  $I^v$ 는 시점  $v$ 에서 취득된 영상들이다.

제안하는 방법에서는 전체 Gaussian  $G_i^{obj}$ 들에 대한 어파인 변환은 Gaussian 들의 평균벡터에 대한 어파인 변환과 같다는 점에 착안하여 수식 (1)의 문제를 단순화시킨다. 즉, 먼저 초기 시점에서 획득한 3D Gaussian 들은 각자의 위치, 방향, 분산 정보와 함께 쉐를 구성한다. 각 Gaussian 의 중심 위치는 평균 벡터(Mean Vector)로 정의되며, 이는 개별 Gaussian 의 공간 좌표를 대표하는 핵심 정보로 사용된다. 이후 객체가 이동하거나 회전하면 두 번째 시점에서 동일한 객체에 대해 새로운 Gaussian 들이 생성되며, 이들 역시 동일한 방식으로 평균 벡터를 계산한다.

그림 2와 같이 Gaussian 간 일관된 비교와 정합을 위해 각 Gaussian 에는 고유 번호가 부여되며,  $P = \{p_1, p_2, \dots, p_n\}$ 와  $Q = \{q_1, q_2, \dots, q_n\}$ 와 같이 시점 간 1:1 대응 관계를 설정할 수 있다. 초기 Gaussian 들의 평균 벡터 집합  $P$ 와 이동 후 Gaussian 들의 평균 벡터 집합  $Q$  간의 차이를 통해 변환 행렬을 추정한다:

3 차원 공간에서 두 점 집합 간의 정합을 위해 OpenCV 의 `cv.estimateAffine3D` 함수를 활용한 3D 아핀 변환 추정 방식을 제안한다. 입력 점 집합  $\{X_i\}$ 와 출력 점 집합  $\{Y_i\}$  간의 관계를 기반으로, 회전 행렬  $R$ 과 이동 벡터  $t$ 로 구성된 아핀 행렬  $A = [R \mid t] \in \mathbb{R}^{3 \times 4}$ 를 최소자승법을 통해 계산한다. 각 점에 대해  $Y_i \approx R \cdot X_i + t$  형태로 정렬되며, 최소 4개 이상의 비공평면 점 쌍이 필요하다.

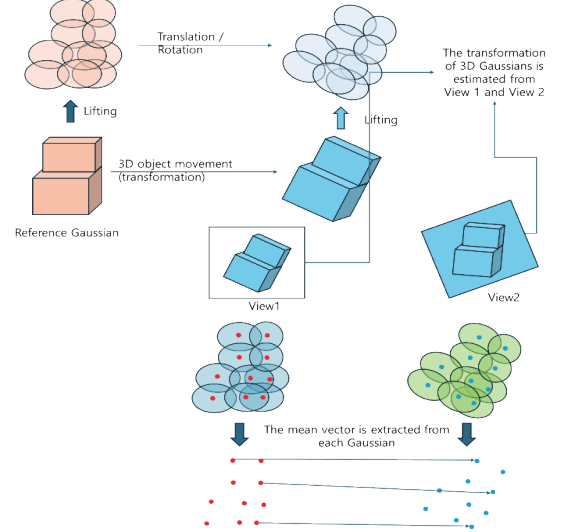


그림 2. 3차원 객체 자세 추정 프로세스

또한, 강인한 추정을 위해 RANSAC 기반 방식이 병행되며, 계산된 자세 파라미터는 기존 Gaussian Scene 에 직접 적용되어 실시간 렌더링에 반영된다. 이 방식은 객체가 지속적으로 움직이는 환경에서도 별도의 학습 없이 효율적인 자세 추정이 가능하며, 3D 영상 및 입체미디어 콘텐츠 제작에 효과적으로 활용될 수 있다.

## III. 결론

본 논문에서는 다른 두 개의 영상으로부터 획득된 Gaussian 들의 평균벡터들 차이를 활용하여 객체의 회전 및 이동 파라미터를 정밀하게 계산하고, 이를 통해 기존 3DGS 쉐의 Gaussian 들을 실시간으로 갱신할 수 있는 방법을 제안한다.

향후 연구에서는 복수 객체 및 복잡한 장면에 대한 확장, 비강체(non-rigid) 객체에 대한 대응, 그리고 텍스처 보존 기반 시각 품질 향상 등의 방향으로 연구를 확장할 예정이다.

## ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [25ZC1110, The research of the basic media · contents technologies]

## 참 고 문 헌

- [1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," arXiv preprint arXiv:2306.00988, 2023.
- [2] B. Smart, C. Zheng, I. Laina, and V. A. Prisacariu, "Splatt3R: Zero-shot Gaussian Splatting from Uncalibrated Image Pairs," arXiv preprint arXiv:2404.07176, Apr. 2024.