

LiDAR-스테레오 융합 기반 3D Gaussian Splatting 최적화 연구

허채연¹, 조영준²

¹전남대학교 인공지능융합학과 석사과정

²전남대학교 인공지능융합학과 교수

cyheo001@jnu.ac.kr, yj.cho@jnu.ac.kr

A Study on 3D Gaussian Splatting Optimization Using LiDAR-Stereo Fusion

Chae-Yeon Heo¹, Yeong-Jun Cho²

¹Dept. of AI Convergence, Chonnam National University

²Dept. of AI Convergence, Chonnam National University

요 약

본 논문에서는 LiDAR 데이터와 스테레오 이미지를 융합하여 고품질 3D 표현을 생성하는 새로운 접근 방식을 제안한다. 제안하는 방법은 LiDAR 포인트 클라우드와 스테레오 비전을 통해 3D 포인트를 추출하는 것에서 시작하며, 이후 스테레오 비전 포인트 클라우드의 스케일을 LiDAR 스케일에 맞추는 조정 과정을 거친다. 스케일이 조정된 스테레오 포인트와 LiDAR 데이터를 초기 융합하여 두 가지 모달리티의 장점을 모두 활용한 포괄적인 포인트 클라우드를 생성한다. 융합된 포인트 클라우드를 정제하기 위해, DBSCAN과 같은 클러스터링을 통한 노이즈 제거와 포인트 그룹화, 그리고 LiDAR 데이터를 기준으로 스테레오에서 추출한 포인트들을 정밀하게 맞추기 위한 회귀 모델을 결합한 하이브리드 기법을 도입한다. 정제된 포인트 클라우드는 3D Gaussian Splatting 초기화를 위한 기초로 사용되며, 각 포인트를 초기 가우시안 값으로 설정하고 다양한 뷰포인트에서의 렌더링 결과를 바탕으로 가우시안 파라미터를 최적화한다. 최적화된 3D 가우시안을 활용하여 다양한 시점에서 장면을 렌더링하고, 이를 통해 연속적이고 풍부한 3D 장면 표현을 생성한다. 본 연구는 일반적인 새로운 뷰 합성(general novel view synthesis) 문제에 대한 중요한 개선을 달성하여, 컴퓨터 비전, 자율주행, 가상현실과 같은 분야에서의 응용 가능성을 보여준다.

1. 서론

자율주행 차량과 고급 로봇 플랫폼 등 자율 시스템의 급속한 발전은 신뢰성 있고 정확한 3D 장면 이해 방법의 필요성을 증가시켰다. 이와 같은 맥락에서 LiDAR 센서와 스테레오 카메라는 상호보완적인 기능으로 인해 널리 사용되고 있다. LiDAR는 정확한 거리 측정과 견고한 3D 포인트 클라우드를 제공하지만, 낮은 해상도와 높은 비용이라는 단점이 있다. 반면, 스테레오 카메라는 고해상도 이미지와 비용 효율성을 제공하지만, 깊이 추정 시 시각적 단서에 의존해야 하는 한계가 있다. 이 두 모달리티의 융합은 각각의 강점을 결합하여 정확하고 세밀한 3D 표현을 생성할 수 있는 가능성을 제공한다. 본 논문은 LiDAR와 스테레오 비전 데이터를 융합하여 정제된 3D 포인트 클라우드를 생성하고, 이를 기반으로 3D Gaussian Splatting(3DGS)[1]을 초기화하고 최적화하는 통합 프레임워크를 제안한다.

제안하는 방법은 먼저 LiDAR와 스테레오 데이터를 활용해 포인트 클라우드를 생성하는 것으로 시작한다. 스테레오 비전 포인트 클라우드는 LiDAR 데이터와 스케일이 다르기 때문에, 두 데이터 소스의 대응점을 찾아 스케일 변환 행렬을 계산하여 스케일을 맞추는 과정이 필요하다. 이 조정 과정은 스테레오에서 얻어진 포인트가 공간적으로 LiDAR 포인트와 일관되도록 하여, 융합된 3D 표현을 가능하게 한다. 스케일 조정 후, 스테레오와 LiDAR 포인트 클라우드를 초기 융합하고, 두 데이터 소스 간의 잡음과 불일치를 줄이기 위해 하이브리드 정제 과정을 적용한다. 이 과정은 DBSCAN과 같은 클러스터링 기법을 사용해 노이즈를 제거하고 유사한 포인트를 그룹화하며, LiDAR 데이터를 기준으로 스테레오 포인트를 정밀하게 정렬하기 위해 회귀 모델을 적용한다. 정제된 포인트 클라우드는 3D Gaussian Splatting[1] 프로세스의 초기화 단계로 활용된다.

3D Gaussian Splatting[1]은 복잡한 장면을 가우

시안 분포의 집합으로 근사하여 표현하는 효과적인 방법이다. 정제된 포인트 클라우드의 각 포인트는 평균, 공분산, 알파 값 등의 파라미터로 초기화된 가우시안으로 표현된다. 이러한 파라미터들은 다양한 시점에서의 렌더링 결과를 바탕으로 손실 함수와 경사 하강법을 사용해 반복적으로 최적화된다. 이 최적화 과정은 3D 가우시안 표현의 정확도를 높이며, 여러 시점에서 포토리얼리스틱한 이미지를 렌더링할 수 있게 한다. 최종적으로, 최적화된 3D 가우시안 표현을 다양한 시점에서 2D 이미지로 렌더링하며, 이를 통해 고품질의 연속적이고 풍부한 3D 장면 표현을 가능하게 한다. 이 접근 방식은 컴퓨터 비전, 자율주행, 가상현실 등 다양한 분야에서의 응용 가능성을 확대한다.

본 논문에서 제안하는 방법은 3D 표현의 품질을 크게 향상시키며, 실험 결과 원본 장면의 세부 사항, 현실감, 충실도가 크게 개선됨을 보여준다. 결론적으로, 본 연구는 LiDAR와 스테레오 데이터를 효과적으로 융합하고, 하이브리드 방식으로 생성된 포인트 클라우드를 정제하여, 3D Gaussian Splatting[1]을 통해 최적화된 장면을 렌더링하는 새로운 프레임워크를 제시한다. 이 프레임워크는 3D 장면 이해와 시각화의 발전을 위한 강력한 도구를 제공하며, 향후 연구 및 다양한 응용에 대한 중요한 시사점을 제시한다.

2. 관련연구

기존의 NeRF[2]와 3D Gaussian Splatting(3DGS)[1] 방법은 다수의 뷰(약 100개)를 입력으로 설계되어, 실제 응용에서 캡처가 번거롭고 비실용적일 수 있다. 이러한 문제를 해결하기 위해 최근에는 소수의 입력 뷰(2개 또는 3개)만으로 장면 재구성과 합성을 수행하는 sparse view 방법이 주목받고 있다. 이러한 sparse view 방법은 크게 두 가지 범주로 나눌 수 있다: 장면별 최적화 방법과 cross-scene feed-forward 추론 방법이다.

장면별 최적화 방법은 최적화 과정을 정밀하게 제어하기 위해 정규화 항을 설계하는 데 중점을 둔다. 이 방법은 개별 장면에 맞춰 최적화를 수행하여 높은 정확도를 보장하지만, 각 장면마다 gradient back-propagation 과정을 반복해야 하므로 시간이 많이 소요되며 실시간 응용에는 비효율적일 수 있다.

반면, feed-forward 모델은 대규모 데이터셋에서

학습된 priors를 사용하여 소수의 뷰만으로 단일 추론을 통해 3D 재구성과 뷰 합성을 수행한다. 이 방법은 장면별 최적화 방법보다 훨씬 빠르고 실시간 응용에 적합하며, 한 번의 추론으로 여러 장면에 일반화할 수 있어 다양한 환경에서 효율적으로 동작할 수 있다. 이는 특히 자율주행과 같은 실시간 데이터 처리 응용에서 중요한 이점으로 작용한다.

3DGS[1]는 NeRF[2]의 높은 계산량 문제를 해결하기 위해 rasterization 기반의 splatting 방식을 사용하여 효율적으로 뷰를 렌더링한다. sparse view를 활용한 3D 재구성을 위해 여러 feed-forward 3DGS 모델이 제안되었으며, SplatterImage[3]와 PixelSplat[4], MVsplat[5] 등이 그 예시이다. SplatterImage[3]는 U-Net을 사용해 단일 뷰에서 가우시안 매개변수를 회귀하지만, 주로 단일 객체에 초점을 맞추고 있어 일반적인 장면 재구성에는 적합하지 않다. PixelSplat[4]은 두 개의 뷰에서 가우시안 매개변수를 회귀하고 epipolar geometry를 사용해 cross-view 인지를 학습하지만, 깊이 추정 시 모호성과 재구성의 한계를 가진다. MVsplat[5]은 멀티뷰 트랜스포머와 2D U-Net을 활용하여 cost volume를 정제하고 깊이 맵을 예측하지만, LiDAR 데이터가 없어 깊이 정보의 정확성에서 한계를 보인다.

3. 제안방법

본 연구에서는 LiDAR 데이터와 스테레오 이미지를 융합하여 정밀한 3D 포인트 클라우드를 생성하고, 이를 바탕으로 3D Gaussian Splatting[1]을 수행하는 새로운 방법을 제안한다. 이 방법은 네 가지 주요 단계로 나뉜다. 데이터 전처리 및 초기 포인트 클라우드 생성, 스테레오 포인트 클라우드의 스케일 조정, 포인트 클라우드 정합 및 병합, 3D Gaussian Splatting[1] 최적화 및 렌더링이다. 각 단계는 3D 장면의 정밀한 표현을 보장하기 위해 설계되었다.

3.1. 데이터 전처리 및 초기 포인트 클라우드 생성

첫 번째 단계는 LiDAR와 스테레오 카메라 간의 캘리브레이션 데이터를 로드하고 초기 포인트 클라우드를 생성하는 것이다. LiDAR와 카메라 간의 상호 캘리브레이션 정보를 포함하는 파일을 읽어들이 변환 행렬과 내·외부 카메라 매트릭스를 추출한다. 예를 들어, 카메라 캘리브레이션 파일로부터 내·외부 파라미터와 LiDAR-카메라 변환 파일로부터 회전 행렬 R 과 변위 T 를 로드하여, 이를 바탕으로

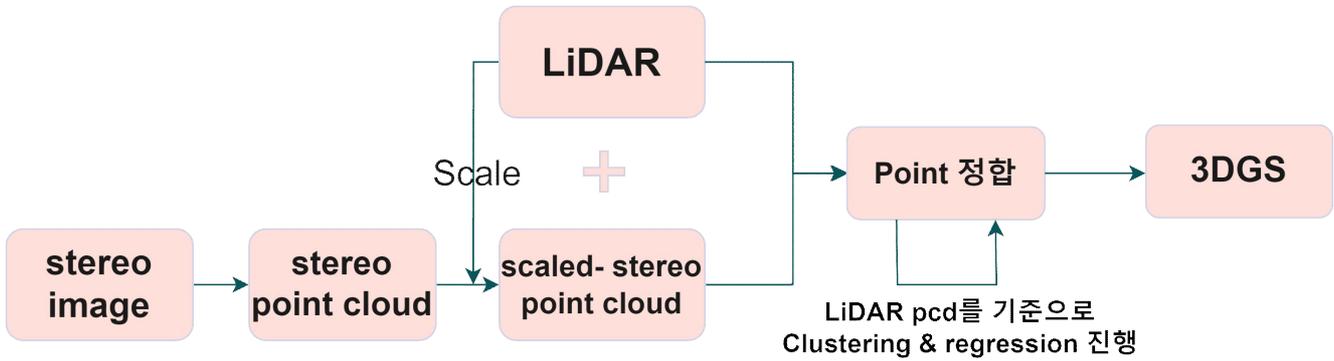


그림 1 . 전체적인 프레임워크

LiDAR와 스테레오 카메라 간의 좌표 변환을 수행한다. 이후, 스테레오 이미지 쌍을 로드하여 3D 포인트 클라우드를 생성하기 위한 기초 데이터를 준비한다. OpenCV의 StereoBM 알고리즘을 사용하여 좌우 이미지 간의 디스패리티 맵을 계산하고, 이 디스패리티 맵은 각 픽셀의 깊이 정보를 포함하며, 이를 활용해 스테레오 이미지로부터 3D 공간 정보를 추출한다.

스테레오 비전의 삼각측량을 위한 Q 매트릭스를 생성하고, 이 매트릭스를 기반으로 디스패리티 맵을 3D 좌표로 변환한다. Q 매트릭스는 카메라의 초점 거리와 기준선 길이(baseline) 등을 포함한 정보를 제공하며, 이를 통해 좌우 이미지 간의 깊이 차이를 3D 공간 좌표로 재구성한다. 이 과정에서 각 픽셀의 3D 좌표와 색상 정보를 포함한 초기 3D 포인트 클라우드가 생성되며, 이 클라우드는 초기 3D 장면 표현의 기초로 사용된다.

3.2.스테레오 포인트 클라우드의 스케일 조정

생성된 스테레오 포인트 클라우드는 LiDAR 포인트 클라우드와 다른 스케일을 가지므로, 두 클라우드의 중심을 맞추고 스케일 차이를 보정하는 작업이 필요하다. 먼저, 디스패리티 맵의 최소값보다 큰 값을 가지는 포인트만을 선택하고, 무한대 값을 가지는 포인트를 제거하여 신뢰할 수 있는 데이터만 남긴다. 그런 다음, 스테레오 포인트 클라우드와 LiDAR 포인트 클라우드의 중심점을 계산하고, 각 포인트에서 중심까지의 평균 거리를 측정하여 스케일 팩터를 산출한다. 스케일 팩터는 스테레오 포인트 클라우드의 크기를 LiDAR 포인트 클라우드에 맞추기 위해 사용되며, 이를 통해 두 데이터 소스 간의 공간적 일관성을 확보한다. 이 조정된 포인트 클라우드는 이후 융합 단계에서 더 정확한 정렬을 보장하게 된다.

3.3.포인트 클라우드 정합 및 병합

이 단계에서는 각각의 포인트 클라우드에 고유한 색상을 부여하여 스테레오 포인트 클라우드는 파란색으로, LiDAR 포인트 클라우드는 빨간색으로 설정한다. 이를 통해 두 데이터 세트를 시각적으로 구분할 수 있으며, 융합 과정에서 각 데이터의 출처를 명확히 할 수 있다. 다음으로, Open3D 라이브러리를 사용하여 포인트 클라우드를 시각화하고, ICP(Iterative Closest Point) 알고리즘을 적용하여 두 클라우드 간의 정밀한 정합을 수행한다. ICP 알고리즘은 스테레오 포인트 클라우드를 LiDAR 포인트 클라우드와 정렬하여 두 클라우드 간의 오차를 최소화하는 변환 행렬을 계산한다. 이를 통해 두 포인트 클라우드가 공간적으로 잘 정렬되도록 하고, 최종적으로 하나의 통합된 포인트 클라우드를 생성한다. 정합이 완료된 후에는 두 포인트 클라우드를 병합하여 하나의 통합된 포인트 클라우드를 생성하고, 이를 파일로 저장한다. 이 병합된 포인트 클라우드는 각 데이터 소스의 강점을 결합하여 더 정확하고 풍부한 3D 표현을 제공한다.

3.4.3D Gaussian Splatting 최적화 및 렌더링

정제된 포인트 클라우드는 3D Gaussian Splatting[1]을 수행하기 위한 초기값으로 사용된다. 각 포인트는 Gaussian 분포의 평균, 공분산, 알파 값 등의 파라미터로 초기화되며, 이러한 파라미터는 포인트의 위치, 밀도, 불확실성 등을 고려하여 설정된다. 이후 다양한 뷰포인트에서의 렌더링 결과를 기반으로 손실 함수와 경사 하강법 알고리즘을 사용하여 Gaussian 파라미터를 반복적으로 최적화한다. 최적화된 Gaussian 파라미터를 사용하여 여러 시점에서 3D 장면을 2D 이미지로 렌더링한다. 이 과정은 연속적이고 현실감 있는 3D 장면을 생성하며, 다양한 응용 분야에서 활용될 수 있다. 최종적으로, 최

적화된 3D Gaussian Splatting[1] 표현은 고품질의 3D 이미지를 생성하여, 자율주행, 가상현실, 컴퓨터 비전 등의 다양한 응용 도메인에서 사용될 수 있다.

이와 같은 방법을 통해 본 연구는 LiDAR와 스테레오 이미지를 효과적으로 융합하여 고품질의 3D 표현을 생성하는 혁신적인 접근 방식을 제시한다. 이 방법은 기존의 시점 합성 문제에 중요한 개선을 이루어냈으며, 실험 결과는 원본 장면에 대한 높은 충실도와 세부 사항을 유지할 수 있음을 입증했다. 이는 다양한 컴퓨터 비전 및 3D 데이터 활용 분야에서 새로운 가능성을 열어주는 중요한 기여를 제공한다.

감사의 글

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었으며 (IITP-2023-RS-2023-00256629), 농림축산식품부의 재원으로 농림식품기술기획평가원의 농식품과학기술융합형연구인력양성사업의 지원을 받아 연구되었음(RS-2024-00397026).

참고문헌

- [1] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4), 2023.
- [2] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *European Conference on Computer Vision (ECCV)*, Virtual, 2020, pp. 405-421.
- [3] Szymanowicz, S., Rupprecht, C., & Vedaldi, A. Splatter Image: Ultra-Fast Single-View 3D Reconstruction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 2024.
- [4] Charatan, David, et al. "pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.2023.
- [5] Chen, Yuedong, et al. "Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images." *arXiv preprint arXiv:2403.14627* (2024).