

# 실시간 Segmentation을 활용한 실내 모바일 로봇 위치 추정 성능 개선

유경민, 남승우, 박재원, 백의준, 김명섭\*

고려대학교

{ rudals2710, nam131119, 2018270614, pb1069, tmskim\* }@korea.ac.kr

## Improving Indoor Mobile Robot Pose Estimation Performance using Real-time Segmentation

Gyeong-Min Yu, Seungwoo Nam, Jae-Won Park, Ui-Jun Baek, Myung-Sup Kim\*

Korea Univ.

### 요약

최근 실내 자율주행 로봇 기술의 수요가 급증하면서, 실내 환경에서 다양한 장애물과 상호작용하며 경로를 효율적으로 계획해야 하는 도전에 직면해 있다. 특히, 벽과 같이 특징점이 명확하지 않은 실내 구조물은 전통적인 VSLAM(Visual Simultaneous Localization and Mapping) 방법으로는 잘 인식되지 않아, 자율주행의 정확성이 저하될 수 있다. 본 연구에서는 이러한 문제를 해결하기 위해 세그멘테이션 기술을 활용하여 벽을 정확하게 분류하고, 이 정보를 바탕으로 로봇의 위치추정 성능을 향상시키는 새로운 방법을 제안한다. 본 논문에서는 grounded segmentation anything 세그멘테이션 기법과 ORB (Oriented FAST and Rotated BRIEF) 기반 특징점 추적 방법을 통합하여, 실내 환경에서 자율주행 성능을 개선한다. 벽과 같이 특징점이 부족한 영역에서 성능 저하 문제를 해결하기 위해, 세그멘테이션을 통해 이러한 영역을 식별하고 특징점을 인공적으로 증강하는 전략을 채택했다. 이 접근 방식을 통해 로봇은 특징점이 부족한 구조물에서도 안정적으로 경로를 추적하고 계획할 수 있다.

### 1. 서론

최근 실내 자율주행 로봇의 수요가 증가하고 있으며, 실내 환경에서 로봇은 다양한 장애물과 상호 작용하며 효율적으로 경로를 계획해야 하며, 이 방법에는 SLAM(Simultaneous Localization and mapping)과 같은 방법이 있다. SLAM 중 VSLAM(Visual Simultaneous Localization and Mapping)은 실내 자율주행 분야에서 중요한 역할을 하는 핵심요소이다. 실내 자율주행 환경의 주요 특징 중 하나는 벽과 같이 질은 질감이 없는 환경이나 조명의 큰 변동이 있는 환경이다. 그러나, 전통적인 특징점 기반 VSLAM 방법에서는 위와 같은 환경에서 특징점을 추정하지 못해 위치 추정 및 전체 VSLAM 시스템의 실패가 발생하는 경우가 많다. 그림 1은 실내 복도를 ORB SLAM[1]을 활용하여 주행한 결과이다. 5번의 실험 모두, 벽과 같이 특징점을 활용해 실내 건물 내 로봇의 위치추정을 시도하게 되면, Tracking Lost가 빈번하게 발생하여 위치 추정이 전체 VSLAM 시스템이 정확하게 이루어지지 않는다는 문제점을 발견하였다. 실내 자율주행 로봇이 효과적으로 주행하기 위해서는 벽과 같은 특징점이 뚜렷하지 않은 구조물을 정확히 인식하고 이를 기반으로 안정적인 경로를 계획할 필요가 있다. 본 논문에서는 이러한 문제를 해결하기 위해 세그멘테이션 기법인 grounded segmentation anything을 활용하여 실시간으로 벽을 정확히 분류하고,

ORB SLAM Tracking state

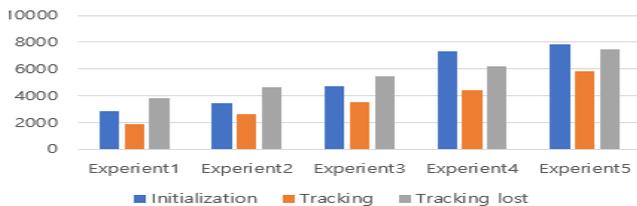


그림 1. 실내 건물 ORB Tracking 시 Tracking State

Fig 1. Indoor ORB Tracking State

세그멘테이션 결과를 특징점으로 사용하여 실내 자율주행 로봇의 위치추정 정확도를 높인다.

### II. 본론

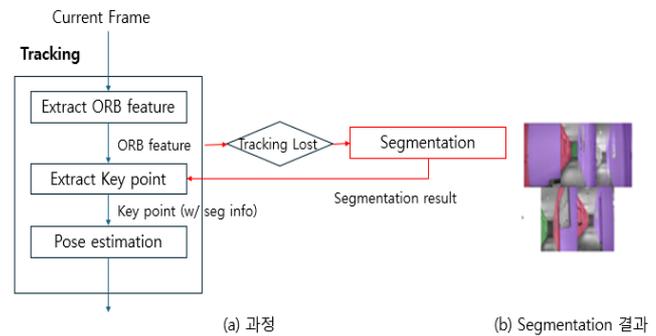


그림 2. (a) Segmentation을 활용한 ORB Tracking System, (b) Segmentation 결과

Fig 2. (a) ORB Tracking System using Segmentation, (b) Segmentation result

[1]과 같은 전통적인 특징점 기반 VSLAM 방법론은 환경 내 뚜렷한 키 포인트 식별에 크게 의존한다. 하지만 이러한 방법들은 벽과 같이 뚜렷한 특징이 부족한 실내 표면에서 종종 추적 실패를 겪는다. 이 때, 의미론적 세그멘테이션(semantic segmentation)은 이미지의 각 픽셀을 사전 정의된 범주(예: 벽, 바닥, 물체 등)로 분류하는 작업으로 특징이 부족한 영역에서의 장애물과 탐색 가능 공간을 식별하는 데 크게 도움을 줄 수 있다. 본 논문은 특징점 기반 VSLAM 시스템의 실내 환경에서의 추적 능력을 개선하기 위해 실시간 세그멘테이션을 활용한 새로운 접근 방식을 제안한다. 본 논문에서는 grounded segmentation anything [2]을 활용하여 실시간으로 실내 환경에서 특징점이 부족한 벽면을 효과적으로 구분하고, ORB 기반 특징점 추적을 통해 로봇의 궤적을 구하는 시스템을 설계하였다. 본 논문의 방법은 그림 2와 같으며, 아래에 설명한다. :

#### 1. 세그멘테이션을 통한 특징점 추적 강화

딥러닝 기반의 세그멘테이션, 특히 grounded segmentation anything[2]

본 논문은 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력 기반 지역혁신 사업(2021RIS-004)과 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0024177, 2023년 지역혁신클러스터육성)을 받아 수행된 연구임.

는 대규모 학습이 진행되어 있어, 별도의 학습 없이 실시간으로 벽과 같은 큰영역을 정확히 구분한다. 이를 통해 실내 환경에서 벽을 식별하고, 세그멘테이션 정보를 활용하여 로봇이 더 효과적으로 주변 환경을 인식할 수 있게 한다. grounded segmentation anything[2]은 Grounding DINO [3] 기술과 SAM[4] 기술이 합쳐진 것으로, Grounding DINO는 제로샷 감지 기능을 갖춘 모델로, 훈련 중에 본 적 없는 객체들을 인식하고 분류할 수 있다. [3]를 활용하여 텍스트를 해석하고, 이미지 내 객체에 대해 정확한 경계 상자 와 레이블을 생성한다. 이 모델은 Vision Transformers (ViTs) 기반이며, 이는 방대한 레이블되지 않은 이미지 데이터셋에서 훈련되어 풍부한 시각적 표현을 학습한다. Segment Anything Model (SAM)은 모든 구별 가능한 객체를 세그먼트할 수 있는 기반 모델이다. 텍스트 묘사뿐만 아니라 경계 상자나 점으로 프롬프트를 처리할 수 있다. 이 세그멘테이션 메커니즘은 카테고리에 관계없이 많은 객체들을 수용한다. SAM은 few-shot 학습 원리를 사용하며 ViTs를 활용하여 다양한 범위의 세그멘테이션 작업에 적용한다. 본 기술의 장점은 학습이 필요없이, 이미 학습되어 있는 모델을 활용하기에 실시간 추적에 용이하다는 점이다.

### 2. 세그멘테이션 영역에서의 특징점 증강

벽이 세그멘테이션되면, 이 영역에 인공적으로 특징점을 무작위로 분포시킨다. 이 방식은 전통적인 특징점 검출기가 텍스처 부족으로 인해 충분한 키포인트를 찾지 못하는 문제를 완화한다. ORB 특징점 추출기를 사용하여 이 증강된 영역에서 특징점을 추출하고, 이를 기반으로 로봇의 이동 경로를 추적한다.

### 3. 증강 특징점을 이용한 궤적 추정

증강된 키포인트를 사용하여 ORB SLAM[1]의 Tracking 부분과 유사한 방식으로 특징점 매칭 및 추적을 수행하지만, 이전에 문제가 되었던 영역에서도 일관된 추적 성능을 보장한다. 매칭된 특징점을 사용하여 로봇의 궤적을 계산하여, 이전보다 더 부드럽고 더 신뢰성 있는 실내 자율주행 로봇의 위치추정을 보장한다.

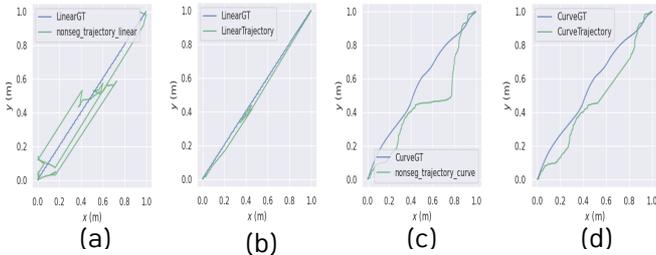


그림 3. (a) 실내 직선 ORB tracking, (b) Segmentation ORB tracking Tracking 결과, (c) 실내 곡선 ORB tracking, (d) Segmentation ORB tracking Tracking 결과

Fig.3 (a) Indoor Linear ORB tracking, (b) Segmentation ORB tracking Tracking result, (c) Indoor Linear ORB tracking, (d) Segmentation ORB tracking Tracking result

	Linear		Curve	
	non seg	seg	non seg	seg
max	0.820223	0.321285	1.305324	1.190046
mean	0.411167	0.213827	0.377106	0.386472
median	0.351432	0.184156	0.286940	0.296762
min	0.288745	0.147786	0.091942	0.0085374
RMSE	0.439504	0.223301	0.455424	0.436644
SSE	8.499208	0.548494	130.046722	77.407273
ssd	0.155259	0.064351	0.0255346	0.203218

표 1 Linear, Curve 구간에서의 Segmentation 유무에 따른 오차값

### 4. 실험적 검증

특징점이 적은 복도에서 직선 및 곡선 주행을 활용해, 기존 ORB SLAM[1]의 Tracking 시스템에 Segmentation을 결합하여 직선 구간과 커브 구간에서 실험을 진행했다. 본 논문의 접근 방식은 특히 복도와 대형 홀과 같이 넓은 균일한 표면으로 인해 특징점 기반 방법에 있어 도전적인 영역에서 추적 실패를 현저히 줄인다. 실험 결과, 세그멘테이션을 수행하지 않았을 때, 그림3에서 볼 수 있듯이, 자신의 위치를 제대로 찾지 못하는 것을 알 수 있다. 그러나, 그림3과 표 1에서 세그멘테이션을 수행한 자신의 위치를 비교적 잘 찾아 Tracking lost가 일어나지 않는 것을 확인할 수 있다. 그러나, 실험 결과, 벽 Segmentation은 잘 이루어지지만, 급격한 시점 변화, 균일 지역에서의 특징 가시성 감소 등의 이유로 곡선 구간에서는 추적 손실을 관찰했다. 추후 연구에서는 특징점을 보완하여 이 문제를 완화하기 위한 제안된 솔루션을 논의할 예정이다.

### III. 결론

본 연구의 목적은 세그멘테이션을 통해 실내 자율주행 로봇이 벽과 바닥을 정확히 인식하고, 이 정보를 활용하여 안정적인 경로 계획을 수행하는 기술을 개발하여 특징점이 뚜렷하지 않은 구조물에서도 효과적인 실내 자율주행이 가능하도록 하며, 실내자율주행 로봇의 위치추정 알고리즘의 성능을 향상시키는 것에 기여하였다. 본 연구에서는 세그멘테이션을 통해 벽을 인식하고, 이를 특징점으로 사용하여 실내 자율주행 로봇의 경로를 계획하는 방법을 성공적으로 개발하였다. 실험 결과, 제안한 방법은 특징점이 뚜렷하지 않은 환경에서도 자신의 위치추정을 성공적으로 수행하였다. 그러나, 모서리가 있는 주행에서는 세그멘테이션 결과를 특징점 증강하는 과정에 대한 고찰이 필요하다. 또한, 본 논문은 특정한 타입의 벽에 대한 세그멘테이션에 최적화되어 있으며, 다양한 텍스처와 조명 조건에서의 성능 향상이 필요하다. 향후 ORB SLAM[1] 전체 시스템과 결합하여 다양하고 복잡한 큰 규모의 실내 구조물에 대해 로봇의 환경 인식을 개선하고, 특징점 분포 방법을 고려해 모서리 추적 능력을 향상시킬 예정이다. 이를 통해 더 복잡한 환경에서의 실내 자율주행 성능을 높이는 방향으로 연구를 확장할 수 있을 것이다.

### 참고 문헌

[1]Mur-Artal, Raul, and Juan D. Tardós. "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras."IEEE transactions on robotics33.5 (2017): 1255-1262.

[2] Liu, Ruiping, et al. "Open scene understanding: Grounded situation recognition meets segment anything for helping people with visual impairments."Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[3]Liu, Shilong, et al. "Grounding dino: Marrying dino with grounded pre-training for open-set object detection."arXiv preprint arXiv:2303.05499(2023).

[4]Ren, Tianhe, et al. "Grounded sam: Assembling open-world models for diverse visual tasks."arXiv preprint arXiv:2401.14159(2024).