

## ARCore SDK 기반 초기 위치 추정을 위한 추론 알고리즘 설계에 관한 연구

조성준, 이강욱, 박주빈, 안선영, 박재영, 윤수연\*  
국민대학교, \*국민대학교 소프트웨어융합대학 소프트웨어학부

kkk121026@kookmin.ac.kr, timsort@kookmin.ac.kr, jbpark0919@kookmin.ac.kr,  
bm9024@kookmin.ac.kr, co2matsign@kookmin.ac.kr, \*1104py@kookmin.ac.kr

Design of an Inference Algorithm for Initial Pose Estimation  
Based on the ARCore SDK

Sung-Jun Jo, Kang-Wook Lee, Jubin Park, Sunyoung Ahn,  
Jaeyoung Park, Soo Yeon Yoon\*  
Kookmin Univ., \*Kookmin Univ.

## 요약

본 논문은 ARCore SDK 기반 위치 데이터를 활용해 모바일 환경에서 실시간으로 동작 가능한 경량 Pose Graph 기반 Visual SLAM 프레임워크를 제안한다. Cross-Track Error(CTE)를 기준으로 성능 실험을 수행한 결과, 평균 CTE는 0.1m 이하로 확인되었고, 특히 Pose Graph 기반 최적화를 적용한 경우, 평균 CTE가 0.0958m로 나타났으며, 미적용 시보다 약 69.4%로 향상된 결과를 확인할 수 있었다. 이는 누적 오차를 효과적으로 보정하고, 궤적의 정확성을 높이는 데 기여함을 의미한다. 최대 CTE 또한 0.3132m로 낮은 오차 범위를 유지하였다. 본 연구는 모바일 기반 실내 내비게이션 및 대규모 공간 데이터 응용에 활용 가능성을 제시한다.

## I. 서론

대형 쇼핑몰, 공항, 병원 같은 복잡한 실내 공간에서 원하는 목적지를 찾지 못해 헤맨 경험은 누구나 한 번쯤 있을 것이다. 안내 표지판은 부족하거나 눈에 잘 띄지 않고, GPS는 건물 내부에선 제대로 작동하지 않는다. 특히 시간에 쫓기거나 처음 방문한 장소라면 이러한 불편은 더욱 크게 다가온다. 이처럼 복잡하고 넓은 실내 공간에서의 정확한 위치 인식과 경로 안내에 대한 수요는 점점 더 커지고 있다. 이러한 문제를 해결하기 위해 로봇 분야에서는 SLAM(Simultaneous Localization and Mapping) 기술이 활발히 사용되어 왔다. 그 중에서도 Visual SLAM은 LiDAR와 같은 레이저 센서 없이 카메라 센서에서 획득한 영상만으로 주변 환경의 지도를 실시간으로 작성하고, 동시에 자신의 위치를 추정할 수 있어, 복잡한 실내 구조에서도 안정적인 주행 및 위치 인식을 가능하게 한다.

그러나 기존 Visual SLAM 알고리즘은 고성능 센서 및 컴퓨팅 자원, 복잡한 시스템 통합을 전제로 설계되어 있어, 배터리 용량과 연산 자원이 제한적인 모바일 디바이스에 바로 이식하기에는 여러 제약이 따른다. 특히, 스마트폰과 태블릿에 내장된 일반 단안 카메라와 IMU만으로도 견고한 Visual SLAM을 구동하기 위해서는 경량화 및 최적화가 필수적이나, 아직 이에 대한 연구는 부족하다.

본 연구에서는 ARCore SDK의 실시간 키프레임 관리 기능과 Pose Graph 최적화 기법을 결합한 모바일 디바이스용 Visual SLAM 프레임워크를 제안한다. 제안하는 방법은 추가 하드웨어 없이 상용 스마트폰만으로도 실내 환경에서 수십 센티미터 수준의 위치 정밀도를 달성하며, 기존 시스템 대비 연산량을 대폭 줄여 처리 시간을 최소화한다.

## II. 관련 연구

## 2.1 Visual-Inertial SLAM

SLAM에는 주로 자율주행 자동차에 설치되는 카메라, LiDAR 와 같은 센서들이 사용된다. 센서로부터 수집한 데이터를 기반으로 자신의 위치를 예측하고 업데이트한다. 그리고 이 과정을 반복함으로써 점차 정확한 지도를 만들어가는 방식으로 SLAM 프로세스가 진행된다. SLAM에 사용되는 센서 중 카메라를 활용한 SLAM 기법을 Visual SLAM 기법이라고 부르며, ORB-SLAM[1-2]과 같이 영상 속 특징점을 기반으로 위치를

추정하는 방식이 대표적으로 사용된다. 하지만 카메라 기반의 SLAM은 빠른 움직임, 조명 변화, 뚜렷한 특징이 없는 환경 등에 취약하다. 이러한 한계를 보완하기 위해 IMU 센서를 결합한 Visual-Inertial SLAM(VI-SLAM) 기술이 주목받고 있다.

IMU는 가속도 및 각속도를 고속으로 측정할 수 있어, 카메라의 프레임 간 시간 차이 동안 발생하는 움직임을 보완해준다. 특히 IMU를 통해 중력 방향, 깊이 정보, 절대적인 자세(roll, pitch) 등을 추정할 수 있어, 단일 카메라 기반 시스템의 깊이 정보 불확실성 문제를 해결하는 데 큰 도움을 준다.

대표적인 VI-SLAM 시스템으로는 Visual-Inertial ORB SLAM[3], ORB-SLAM3[4], VINS-Mono[5] 등이 있으며, 이들은 실시간 루프 클로징 및 지도 재사용 기능을 통해 정밀한 위치 추정을 실현하고 있다.

## 2.2 ARCore

ARCore는 구글이 2018년에 발표한 모바일 증강현실(AR) 플랫폼이다. ARCore의 핵심 모듈 중 하나인 Motion Tracking 모듈은 Visual-Inertial SLAM 기법을 기반으로, 카메라 영상에서 추출한 특징점과 IMU 데이터를 융합하여 디바이스의 초기 위치를 빠르게 계산한다.[6] 그러나 이 방식은 절대 위치 추정을 지원하지 않기 때문에, 위치 정보는 오직 상대적인 기준에 의존하게 된다. 이로 인해 시간이 지남에 따라 오차가 누적되며, 정확도가 저하된다. 따라서 내비게이션과 같이 장시간 위치를 추정하는 환경에는 적합하지 않다.

본 연구에서는 ARCore의 Motion Tracking 모듈에 Pose Graph 자료 구조를 결합하여, 현재 위치와 현재와 매칭되는 과거 위치를 가져오고, 두 위치 간의 변위 및 관측값 간 변위 차이를 여러 값으로 정의하여 이를 최소화하는 방식으로 최적화를 수행한다. 이러한 접근을 통해 로봇용 Visual SLAM 수준에 준하는 고정밀 위치 추정 기술을 모바일 디바이스 상에서도 구현하고자 한다.

## III. 실험 및 결과 분석

## 3.1 Pose Graph Optimization

본 연구는 모바일 디바이스 상에서 실시간 동작 가능한 Visual SLAM 시스템 구현을 목표로 하였으나, 초기 구현에서는 Pose Graph Optimization 단계에서 높은 연산 부하로 인해 CPU 사용량의 급격한 증

가, 디바이스 과열, 프레임 드랍이 심화되는 현상이 나타났다. 특히, 최적화 과정에서 반복적으로 수행되는 자동 미분 (auto-differentiation)이 주요한 병목으로 작용하였다.

위와 같은 문제를 해결하기 위해, 본 연구에서는 Ceres의 Costfunction을 오버로딩해 자동 미분을 수동 미분(manual differentiation) 방식으로 대체하였다. 각 노드 간 상대 pose에 대해 정의된 잔차(residual)에 대한 Jacobian을 분석적으로 도출하고, 이를 Ceres Solver의 최적화 과정에 명시적으로 제공해서 자동 미분에 소요되는 연산을 제거했다. 그 결과, 최적화 과정에서의 연산 오버헤드가 크게 감소했다. 이로 인해 CPU 사용률이 안정적으로 유지되었고, 디바이스 과열이나 프레임 지연 현상이 해소되었다.

실험 결과는 [표 1]과 같다. 동일한 최적화 문제에 수동 미분 방식과 자동 미분 방식을 비교한 결과, 최적화 루프당 소요 시간은 약 89.4% 향상되었다. 더불어 모바일 디바이스에서의 연산 부하가 경감됨에 따라 SLAM 성능이 개선되었으며 이로 인해 실시간 동작이 가능해졌다.

최적화 방식	평균 처리 시간 (ns)	시간 차이 (ns)	처리 시간 비교
수동 미분	40,260	341,415	수동 미분이 89.4% 단축
자동 미분	381,675		

[표 1] 최적화 방식에 따른 소요시간 비교

### 3.2 실험 방법

본 연구에서는 ARCore와 Pose Graph를 결합한 Visual SLAM 시스템의 위치 추정 성능을 검증하고자 실제 환경에서 보행 기반 실험을 수행하였다. 실험은 국민대학교 미래관 6층 복도에서 진행되었으며, 약 1m 간격으로 정밀 측정된 기준 위치(Ground Truth, GT)를 바닥에 표기하여 정답 경로로 활용하였다.

Pose Graph 기반 궤적 최적화의 효과를 평가하기 위해 동일한 경로를 두 가지 조건에서 각각 보행하였다. 첫 번째 조건은 Pose Graph 최적화 모듈이 포함된 버전이며, 두 번째 조건은 해당 모듈을 제외한 버전이다. 두 실험 모두 동일한 모바일 디바이스를 사용하여 실시간으로 위치를 추정하고 그 결과를 저장하였다. 저장된 Pose에 대한 그래프는 [그림 1]에서 확인할 수 있는데, Pose Graph 최적화 모듈이 포함된 버전이 보다 GT에 가까운 것을 볼 수 있다.



[그림 1] 경로 최적화 전후 결과와 정답 경로 시각화

### 3.3 실험 결과 분석

시스템 성능 평가는 시간대별 추정 위치와 이에 대응하는 정답 위치 간의 거리인 Cross-Track Error (CTE)를 기준으로 수행하였다. 그 결과, 제안 시스템은 평균 CTE 0.1 m 이내, 최대 CTE 0.3 m 수준의 오차를 기록하여, 실내 환경에서도 안정적이고 정밀한 위치 추정 성능을 보이는 것으로 나타났다.

특히, [표 2]를 참고하면, 제안된 시스템인 Pose Graph 기반 최적화가 적용된 경우의 평균 CTE는 0.0958 m로, 최적화를 생략한 경우의 평균 CTE인 0.3132 m에 비해 약 69.4% 감소하였다. 이는 실시간 위치 추정 과정에서 누적 오차가 점진적으로 증가하는 경향이 있음을 시사하며,

Pose Graph 최적화 기법이 이러한 오차를 효과적으로 보정하여 전체 궤적의 정확성을 향상시킨다는 것을 실험적으로 입증한다.

최적화 여부	평균 CTE (m)	최대 CTE (m)	평균 CTE 비교
Pose Graph 적용	0.0958	0.3132	Pose Graph 적용 시 69.4% 향상
Pose Graph 미적용	0.3232	0.8675	

[표 2] Pose 최적화 여부에 따른 CTE

## IV. 결론

본 연구에서는 ARCore 기반으로 획득한 위치 데이터를 활용하여 독자적인 Pose Graph를 구성하고, 이를 모바일 환경에서도 효율적으로 운용할 수 있도록 경량화 방법을 제안한다. 기존의 Pose Graph 기반 SLAM 기법들이 고성능 컴퓨팅 자원을 필요로 했던 것과 달리, 본 논문에서는 연산량 최적화 및 데이터 구조의 간소화를 통해 모바일 디바이스에서도 실시간 운용이 가능한 수준으로 기술을 경량화하였다.

이를 통해 ARCore의 Motion Tracking 기능을 넘어, 공간 이해 및 최적화된 경로 복원이 가능하다. 더불어 최적화된 Pose Graph 구성 및 업데이트를 통해 기존에 모바일 디바이스에서는 적용이 어려웠던 SLAM을 사용할 수 있다.

결론적으로, 본 연구는 ARCore 기반의 위치 데이터를 활용하여 이를 모바일 환경에 맞춰 사용할 수 있는 접근법을 제안한다. 나아가 백화점, 박물관과 같은 넓은 실내 공간에 적용하여 다양한 사용자에게 높은 정확도의 위치 인식과 공간 이해를 제공할 수 있다. 이는 사용자 맞춤형 내비게이션, 실시간 정보 제공, 군중 흐름 분석 등 다양한 응용 서비스로 확장될 수 있으며, 특히 별도의 고성능 장비 없이 모바일 디바이스만으로 정밀한 공간 서비스를 실현할 수 있다는 점에서 큰 의의가 있다.

## ACKNOWLEDGMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음.(2022-0-00964)

"This research was supported by the MIST(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation) in 2022"(2022-0-00964)

## 참 고 문 헌

- [1] Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D. Tardos. "ORB-SLAM: a versatile and accurate monocular SLAM system." IEEE transactions on robotics 31.5 (2015): 1147-1163.
- [2] Mur-Artal, Raul, and Juan D. Tardos. "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras." IEEE transactions on robotics 33.5 (2017): 1255-1262.
- [3] Mur-Artal, Raúl, and Juan D. Tardós. "Visual-inertial monocular SLAM with map reuse." IEEE Robotics and Automation Letters 2.2 (2017): 796-803.
- [4] Campos, Carlos, et al. "Orb-slam3: An accurate open-source library for visual, visual - inertial, and multimap slam." IEEE Transactions on Robotics 37.6 (2021): 1874-1890.
- [5] Qin, Tong, Peiliang Li, and Shaojie Shen. "Vins-mono: A robust and versatile monocular visual-inertial state estimator." IEEE transactions on robotics 34.4 (2018): 1004-1020.
- [6] <https://developers.google.com/ar/develop/fundamentals>