

Visual-Inertial SLAM 기반 실내 위치 및 환경 인식 시스템 설계에 관한 연구

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

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

Design of an Indoor Localization and Environment Perception System Using Visual-Inertial SLAM

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

요약

본 논문은 ARCore 기반 상대 위치 추정에 VINS-Mono 기반 Pose Graph Optimization을 결합하여, 모바일 디바이스만으로 실시간 실내 내비게이션이 가능한 경량화된 시스템을 제안한다. ARCore의 상대 이동 정보를 연속적으로 최적화함으로써 누적 오차를 줄이고 절대 위치를 정밀하게 추정한다. 국민대학교 미래관 4층에서의 실험에서는 STT 기반 LLM 음성 인식을 통해 목적지를 설정하고, A* 알고리즘을 통해 경로를 탐색 및 TTS로 안내하였다. STT 인식 성능은 총 200회의 테스트에서 평균 96%의 정확도를 기록하였으며, 이는 사용자의 목적지를 정확히 인식하여 경로 탐색의 안정성을 높이는 데 기여하였다. 경로 안내의 평균 CTE 오차는 0.419m로, 실내 환경에서도 우수한 위치 정확도를 입증하였다. 본 연구는 외부 센서나 인프라 없이도 저가 모바일 기기만으로 실내 위치 인식과 경로 안내가 가능함을 실증하였다.

I. 서론

내비게이션은 모바일 디바이스, 차량, 로봇 등 다양한 영역에서 필수적인 기능이다. 실외 환경에서는 Global Positioning System (GPS)를 기반으로 한 위치 추정 기술이 발전하며 높은 정확도로 사용자에게 안정적인 경로 안내 서비스를 제공하고 있다. 그러나 GPS는 위성을 기반으로 신호를 수신하는 특성상, 실내에서는 벽이나 천장 같은 구조물에 의해 신호가 차단되거나 왜곡되어 정확한 위치 파악이 어렵다. 이로 인해 대형 건물, 학교, 병원 등에서의 길찾기는 큰 불편으로 남아 있다. 이러한 문제를 해결하기 위해 Wi-Fi, BLE와 같은 다양한 실내 내비게이션 기술이 제안되어 왔으나, 대부분 별도의 인프라 설치가 필요하거나 환경 제약이 커서 모든 공간에 범용적으로 적용하기에는 한계가 존재한다.

이에 본 연구에서는 로봇에서 자주 사용하는 Visual-Inertial Simultaneous Localization and Mapping (Visual-Inertial SLAM) 기술을 활용하여 별도의 초기 비용 없이 모바일 디바이스에서 사용할 수 있는 실내 내비게이션 애플리케이션 개발 방안을 제시한다.

II. 관련 연구

2.1 Visual-Inertial SLAM

Visual SLAM은 카메라 센서로만 위치를 추정하기 때문에 빠른 움직임, 조명 변화, 텍스처가 부족한 환경과 같이 시각적 특징 추출이 어려운 상황에서 성능이 저하되는 한계를 지닌다. 이러한 문제를 보완하기 위해 IMU 센서를 결합한 VI-SLAM 기술이 주목받고 있다.

VI-SLAM은 카메라 영상에서 추출한 시각적 특징과 IMU로부터 얻은 관성 데이터를 정밀하게 융합하여, 로봇의 위치와 자세를 추정하는 기술이다. 이 방식은 급격한 움직임에 의해 발생하는 오차를 IMU 데이터를 통해 보정하며, IMU가 제공하는 기기의 기울기(roll, pitch) 정보와 실제 크기(scale)에 대한 추정을 기반으로 한다. 이렇게 예측된 위치 및 자세는 시각적 특징을 활용하여 주기적으로 보정된다.

대표적인 사례로, Visual-inertial monocular SLAM[1], ORB-SLAM3[2], VINS-Mono[3] 등이 있으며, 이들은 실시간 루프 클로징 및 지도 재사용 기능을 통해 높은 정밀도의 위치 추정을 구현한다.

특히, VINS-Mono는 위치(x, y, z)와 방향(yaw)에 해당하는 4 자유도만을 최적화하는 Pose Graph Optimization 기법을 적용함으로써, 연산 효율성이 뛰어나다.

2.2 ARCore

ARCore는 구글이 2018년에 발표한 모바일 증강현실(AR) 플랫폼이다. ARCore의 핵심 모듈 중 하나인 Motion Tracking 모듈은 VI-SLAM 기법을 기반으로, 카메라 영상에서 추출한 특징점과 IMU 데이터를 융합하여 디바이스의 초기 위치를 빠르게 계산한다.[4] 특히 ARCore는 카메라의 내부 파라미터 (주점, 초점 거리 등)를 자동으로 추정하여 제공하므로, 사용자나 개발자가 별도로 기기의 파라미터를 수동으로 보정하지 않아도 바로 상대 위치를 추정할 수 있다는 장점이 있다. 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 VINS-Mono의 Pose Graph Optimization 기법을 결합하여, 현재 위치와 현재와 매칭되는 과거 위치를 가져오고, 두 위치 간의 변위 및 관측값 간 변위 차이를 에러값으로 정의하여 이를 최소화하는 방식으로 최적화를 수행한다. 이러한 접근을 통해 로봇용 VI-SLAM 수준에 준하는 고정밀 위치 추정 기술을 모바일 디바이스 상에서도 구현하고자 한다.

III. 실험 및 결과 분석

3.1 실험 설계

3.1.1 ARCore 기반 위치 추정 및 VI-SLAM 맵 생성

ARCore는 모바일 디바이스에 내장된 카메라와 IMU 센서를 활용하여 Visual-Inertial Odometry를 통해 기기의 상대적 위치 및 자세를 실시간으로 추정하고 있다. 하지만 루프 클로징(loop closure)나 누적 오차(drift)에 대한 보정 기능은 제공하지 않기 때문에, 본 연구에서는 VINS-Mono 기반 VI-SLAM 기술을 도입하여 ARCore로부터 획득한 위치 정보를 Pose Graph 형태로 저장한다. 저장된 그래프는 후처리 단계

에서 최적화 과정을 거쳐 누적 오차를 보정한 고정밀 사전 맵으로 재구성되며, 이를 실시간 위치 추정의 기준 맵으로 활용하고 있다.

3.1.2 A* (A-star) 알고리즘 기반 경로 탐색

ARCore 기반의 실시간 위치 추정이 수행된 후, 사용자의 위치를 출발점으로 하여 목적지까지의 최적 경로 탐색을 위해 A* 알고리즘을 적용하였다. A* 알고리즘의 주요 동작 방식은 다음과 같다.

- ① 노드 확장: 현재 위치에서 8방향으로 이동 가능한 다음 노드를 생성하고, 각 노드의 g, h, f 값을 계산하여 우선순위 큐(priority queue)에 삽입한다.
- ② 휴리스틱 함수: 목표 지점까지의 예상 거리는 유클리드 거리($\sqrt{(x_2 - x_1)^2 + (z_2 - z_1)^2}$)를 사용하여 계산된다.
- ③ 경로 종료 조건: 현재 노드가 목표 지점으로부터 일정 거리 (0.8m 이하)에 도달하면 탐색을 종료하고 경로를 구성한다.
- ④ 경로 재탐색: 경로를 따라 이동 중 사용자의 위치가 허용 편차 (2.0m)를 초과하여 벗어날 경우, 경로 이탈로 판단하고 현재 위치를 기준으로 새로운 경로를 탐색한다.

3.1.3 음성 기반 경로 입력 및 안내 시스템

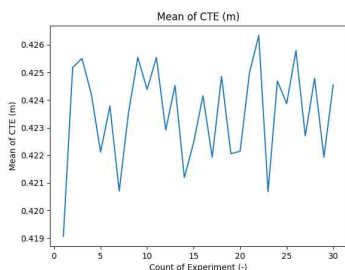
노약자나 시각장애인 등 다양한 사용자의 접근성을 고려하여, 본 연구에서는 음성 기반 인터페이스를 설계하였다. 이 시스템은 경로 안내용 음성 출력과 사용자의 명령 인식을 위한 음성 입력 기능으로 구성된다. Text-to-Speech (TTS) 기능은 Google의 TTS API를 기반으로 구현되었으며, 사용자의 이동에 따른 이벤트에 따라 음성 안내를 제공한다. 이러한 음성 기반 인터페이스는 화면 조작이 어려운 사용자도 손쉽게 실내 내비게이션 기능을 활용할 수 있도록 지원하며, 접근성과 편의성 모두를 향상시키는 핵심 요소로 작용한다.



[그림 1] WIGO 사용자 인터페이스(UI)

3.2 인터페이스

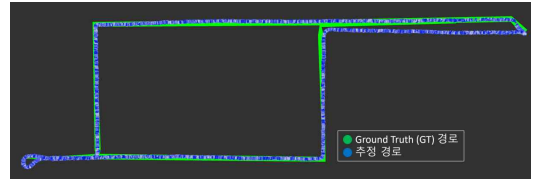
[그림 1]는 WIGO 애플리케이션의 실시간 경로 안내 화면이다. 가장 바깥쪽 원에 위치한 WIGO 아이콘은 사용자가 현재 바라보는 방향을 나타낸다. 그 아래에 표시된 빨간 점은 A* 알고리즘으로 생성된 경로 중 다음 경로 방향을 가리키며, 사용자의 방향과 이 목표 방향이 일치하면 진동 알림이 발생한다. 바닥에 연속적으로 표시된 흰색 화살표는 A* 알고리즘에 의해 생성된 경로 노드이다. 상단에는 목적지 검색 창이 있으며, 우측 상단의 톱니바퀴 아이콘을 통해 환경 설정 및 지도 관리 기능에 접근할 수 있다. 또한, 음성 아이콘을 통해 음성 명령을 사용할 수 있다.



[그림 2] 실험 환경에 대한 그래프

3.3 실험 결과 분석

실험 결과, 전체 경로 길이 약 300m 기준으로 실시간 위치 추정의 평균 오차는 0.4191m 수준으로 측정되었다. 경로 탐색 및 안내 정확도는 총 30회의 반복 실험에서 모두 사용자를 정확히 목표 지점까지 유도하는데 성공하였으며, [그림 2]에 나타나있다. 특히, AR 기반 인터페이스 상에서 사용자 시선 방향과 경로 방향이 시각적으로 명확히 구분되어, 사용자가 올바른 방향으로 이동하는 데 큰 도움이 되는 것으로 분석되었다.



[그림 3] Ground Truth 경로 및 직접 취득한 경로

[그림 3]는 국민대학교 미래관 4층에 대한 전체 약 300m의 경로를 시각화한 것이다. 녹색으로 표시된 부분은 실측한 GT이고, 청색으로 표시된 부분은 우리 애플리케이션을 바탕으로 측정한 경로이다. 그림에서 확인할 수 있듯, 긴 경로에서도 예측 경로는 정답 경로와 큰 차이 없이 잘 일치하는 것을 볼 수 있다. 이를 통해 매우 긴 경로에서도 정확한 경로 안내가 가능할 것으로 기대된다.

IV. 결론 및 시사점

연구에서는 ARCore의 위치 추정 기능과 Pose Graph 기반의 최적화 기법을 결합하여, 모바일 디바이스 환경에서도 실시간으로 동작 가능한 VI-SLAM 기반의 실내 내비게이션 시스템을 제안한다. 이를 사용하여 저가의 상용 기기에서 별도의 외부 센서나 인프라없이 정확한 위치 인식과 경로 안내가 가능하다.

제안한 시스템은 단순한 위치 추적을 넘어, 다양한 응용 분야로의 확장 가능성을 확인할 수 있었다. 박물관이나 전시 공간에서는 AR 기반 큐레이션 시스템으로 활용되어 관람객의 위치에 따라 맞춤형 전시 정보를 제공하여 사용자 경험을 높일 수 있으며, 대형 쇼핑몰이나 백화점 같이 길을 찾기에 복잡한 구조의 건물에서 사용자 맞춤형 경로 안내를 제공하는 데 기여할 수 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)

V. 참고문헌

- [1] Mur-Artal, Raúl, and Juan D. Tardós. "Visual-inertial monocular SLAM with map reuse." IEEE Robotics and Automation Letters 2.2 (2017): 796-803.
- [2] 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.
- [3] 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.
- [4] <https://developers.google.com/ar/develop/fundamentals>