엣지 컴퓨팅 기반 단안 깊이 추정 모듈과 VSLAM 통합 시스템 설계

유경민, 남승우, 박재원, 백의준, 김지민, 김명섭*

고려대학교, *고려대학교

rudals2710@korea.ac.kr, nam131119@korea.ac.kr, 2018270614@korea.ac.kr, pb1069@korea.ac.kr, illiard1209@korea.ac.kr, *tmskim@korea.ac.kr

Design of an Edge Computing-Based Monocular Depth Estimation Module and VSLAM Integrated System

Yu Gyeong Min, Nam Seung Woo, Park Jae Won, Baek Ui Jun, Kim Ji Min, Kim Myeong Sub*

Korea Univ., Korea Univ., Korea Univ., Korea Univ., Korea Univ.,*Korea Univ.

요 약

본 논문은 엣지 디바이스에서 ORB-SLAM2를 실행하고, 클라우드 서버에서 Depth Anything v2를 병렬로 수행하는 분산형 VSLAM 아키텍처를 제안한다. 제안된 구조는 GPU가 없는 저전력 로봇 환경에서도 고정밀의 단안 깊이 추정을 가능하게 하며, 전체 SLAM 시스템의 정밀도와 효율을 동시에 향상시킨다.TUM RGB-D 데이터셋 기반 실험 결과, ORB-SLAM2 단독 사용 시 평균 절대 궤적 오차(ATE)는 0.442m였으나, 제안한 구조에서는 0.024m로 94.6% 감소하였다. 또한, 저조도 환경(freiburg3_nostructure_texture_far)이나 텍스처가 부족한 장면에서도 ATE가 0.041m에서 0.029m로 감소하여, 다양한 조건에서도 높은 안정성과 정밀도를 유지함을 확인하였다. 본 구조는 자율주행, 실내 내비게이션 등 실시간성과 정밀성이 요구되는 응용 분야에 실용적인 솔루션으로 적용 가능함을 입증한다.

Ⅰ. 서 론

비전 기반 동시 위치 추정 및 지도 작성(VSLAM, Visual Simultaneous Localization and Mapping)은 카메라 센서를 이용하여 주변 환경을 인식하고, 이동체의 위치를 추정하며, 3차원 지도를 구축하는 핵심 기술로, 자율주행, 로봇 내비게이션, 증강현실(AR) 등 다양한 응용 분야에서 활용된다. 특히, 단일 카메라만을 사용하는 단안 VSLAM(Monocular VSLAM)은 하드웨어 구성의 단순성과 비용 효율성 측면에서 장점을 가지며, 소형 로봇이나 저전력 디바이스 기반 플랫폼에 적합한 구조로 주목받고 있다. 그러나 단안 VSLAM은 깊이 정보를 직접적으로 측정할 수 없기 때문에 스케일 모호성(Scale Ambiguity)과 누적 오차(Scale Drift) 문제가 발생하며, 이로 인해 지도 왜곡 및 위치 추정 정확도의 저하(Scale Drift)가 발생하는 한계를 지닌다. ORB-SLAM2 [1]와 같은 대표적인 단안 VSLAM 시스템에서도 이러한 문제가 반복적으로 보고되고 있다.

이를 보완하기 위해 최근에는 단안 이미지 기반의 딥러닝 기반 깊이 추정 기법들이 제안되고 있으나, 대부분의 최신 모델들은 고성능 GPU 자원을 필요로 하며, 메모리 및 연산량이 크기 때문에 엣지 환경에서는 실시간 적용이 어렵다.

본 논문에서는 ORB-SLAM2[1]를 엣지 디바이스에서 실행하고, 선택된 KeyFrame만 클라우드로 전송하여 Depth Anything v2[2] 기반 단안 깊이 추정을 수행하는 분산형 협업 구조를 제안한다. 이 구조는 연산 자원을 분산하고 통신 효율을 극대화하며, 엣지 기반 저전력 로봇 시스템에서도 높은 정확도의 실시간 VSLAM 구현을 가능하게 한다.

Ⅱ. 본론

Table 1. 시스템 구성 및 사양

구성요소	사양	역할
엣지장치	NVIDIA Jetson AGX Orin (32GB), 8-core ARM CPU	VSLAM (ORB-SLAM2) 실행
로봇센서	단안 카메라(30fps)	환경 데이터 수집
클라우드 서버	NVIDIA GeForce RTX 4090 x4 (CUDA 12.2)	단안 깊이 추정 (Depth Anything v2)

2.1 전체 아키텍처 개요

Figure 1은 제안하는 시스템의 전체 구조를 보여준다. 본 시스템은 엣지 기반 ORB-SLAM2[1] 처리 모듈, 클라우드 기반 단안 깊이 추정 모듈, 그리고 양방향 통신 및 통합 최적화 모듈의 세 가지 주요 구성 요소로 이루어져 있다. ORB-SLAM2[1]는 엣지 디바이스에서 실시간으로 동작하며, 시스템 내부에서 선택된 주요 KeyFrame만이 클라우드로 전송된다. 클라우드에서는 Depth Anything v2[2]를 통해 정밀한 깊이 맵을 생성하고, 이를 다시 엣지로 전송하여 SLAM 시스템에 통합하게 된다. 이러한 구조는 연산 부하를 엣지와 클라우드 간에 효과적으로 분산시키고, 통신 대역폭을 절감하는 동시에, 높은 정확도와 실시간성을 모두 확보할 수 있는 장점을 갖는다.

2.2 엣지 기반 ORB-SLAM2 처리 모듈

엣지 단에서는 NVIDIA Jetson AGX Orin과 같은 ARM 기반 장치를 사용하여 ORB 특징점의 추출 및 추적, KeyFrame 선별과 전송, 그리고 초기화 및 로컬 맵 관리 등의 기능을 수행한다. 전체 680프레임 중 약

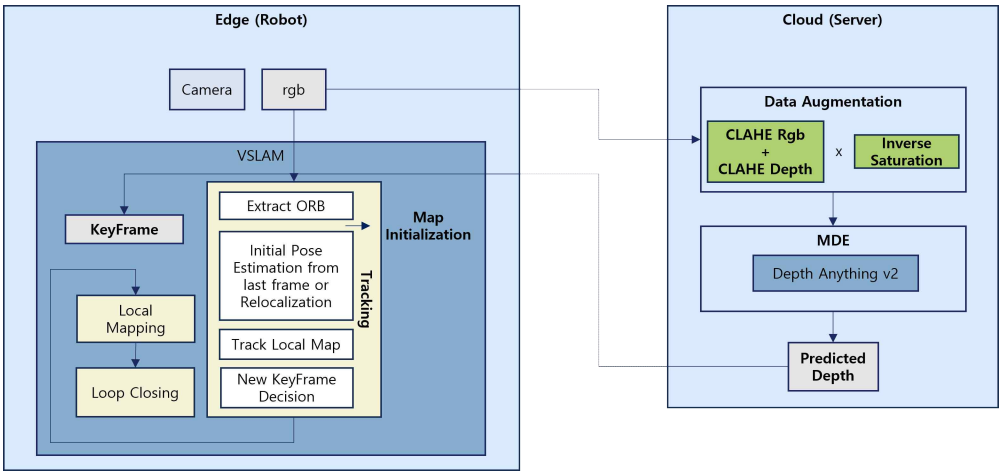

Fig 1. System Architecture

15~20%만이 KeyFrame으로 클라우드에 전송되며, 초기화에 사용되는 프레임은 약 8개로 제한되어 통신량은 평균적으로 95% 이상 절감된다. 클라우드에서 반환된 깊이 정보는 ORB-SLAM2[1]의 초기화 및 로컬 맵 생성 과정에 통합되어 스케일 정합도를 높이고 누적 오차를 줄이는 데 기여한다.

2.3 클라우드 기반 단안 깊이 추정 모듈

클라우드 서버는 NVIDIA RTX 4090 GPU 네 개를 활용하여 고성능 연산을 수행한다. 클라우드는 엣지에서 전송된 KeyFrame을 수신하고 전처리하며, Depth Anything v2[2]를 활용해 단안 영상으로부터 깊이 맵을 정밀하게 추정한 후 이를 다시 엣지로 전송한다. 클라우드 측에서 생성된 깊이 맵은 ORB-SLAM2[1]의 KeyFrame에 통합되며, 이로 인해 맵포인트 초기화 및 최적화 과정에서 더욱 정밀한 깊이 정보를 제공하게 된다.

Table 2 Result of Our System

	Non cloud	Edge Cloud Conmputing			
항목	non	All	Initial	KeyFrame	ours
ATE(m)	0.044	0.19	0.10	0.11	**0.024**
통신 데이터량	0	680	8	57	65
실시간성 (FPS)	34.8	2.5	**28.5**	21.2	18.8
초 기 화 실패율	높음	높음	중간	높음	**낮음**

2.4 양방향 통신 및 최적화 전략

본 시스템은 실시간 처리를 유지하기 위해 다양한 전략을 적용하였다. 첫째, 모든 프레임이 아닌 의미 있는 프레임만을 선택적으로 클라우드로 전송함으로써 통신 지연과 데이터 전송량을 최소화하였다. 둘째, 클라우드에서 반환된 깊이 정보는 ORB-SLAM2[1]의 맵포인트 생성, Bundle Adjustment, 루프 클로징 등 다양한 처리 단계에 통합되어 전체적인 정밀도 향상에 기여한다.

비교 실험 결과는 Table 2에 정리되어 있다. 비교 대상에는 Non-Cloud 방식, 모든 프레임을 클라우드로 전송하는 방식(All), 초기화 프레임만 전송하는 방식(Initial), KeyFrame만 전송하는 방식(KeyFrame), 그리고 제안 방식(Ours)이 포함된다. 절대 궤적 오차(ATE)는 제안 방식이 0.024m로 가장 낮았으며, 통신 데이터량은 전체 프레임 중 약 9.5%인 65프레임만 전송되어 효율적이다. 실시간성(FPS) 측면에서도 제안 방식은 평균 18.8 FPS로, 640×480 해상도 기준 실제 응용 환경에서의 실시간 처리를 만족시킨다. 초기화 실패율 역시 제안 방식이 가장 낮게 나타났다.

2.5 성능 비교 요약

본 시스템의 성능은 TUM RGB-D 벤치마크[3]를 기반으로 기존 ORB-SLAM2[1]와 비교되었다. 이 벤치마크는 휴대용 Kinect 카메라로 촬영된 컬러 영상과 깊이 영상을 포함하며, 30Hz 주기로 수집된 640×480 해상도의 이미지 시퀀스를 제공한다. 카메라의 실제 이동 경로는 고정밀 모션 캡처 시스템을 통해 100Hz로 수집되어, SLAM 시스템의 위치 정확도를 정량적으로 평가할 수 있다. 본 연구에서는 예측 궤적과 실제 궤적 간의 절대 거리 차이를 나타내는 ATE를 주요 평가 지표로 사용하였다. 모든 실험은 VSLAM의 무작위 초기화 및 자세 최적화 특성을 고려하여 5회 반복 수행되었고, 평균값을 기준으로 평가하였다. 오차가 가장 적은 결과는 볼드체로 표기하였다.

실험 결과, 모든 프레임을 클라우드로 전송하는 방식은 ATE 0.19m, FPS 2.5로 실시간성은 매우 낮았고, 초기화 프레임만 처리하는 경우 FPS는 28.5로 높았지만 ATE는 0.10m로 정확도는 낮았다. 반면 제안하는 방식은 ATE 0.024m, FPS 18.8로 정확성과 실시간성을 모두 만족하며, 통신 프레임 수도 전체의 약 9.5%로 가장 효율적인 결과를 나타냈다. 또한 생성된 지도는 Figure 2에서 확인할 수 있듯, 단안 카메라만 사용하는 기존 방식보다 훨씬 정확한 결과를 제공하였다.

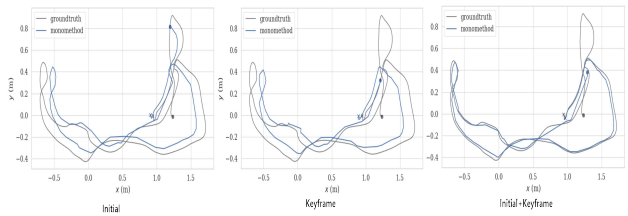

Fig. 2. ORB-SLAM3 MDE with initial,, KeyFrame, Our System(Initial+KeyFrame)

Table 3. 저텍스처 구조에서의 성능

	Mono	
	ORB SLAM [1]	Ours
fr3_nt_t_f	0.853m	**0.044m**
fr3_nt_t_n	0.473m	**0.045m**
fr3_t_nt_f	0.904m	**0.051m**
fr3_t_nt_n	0.693m	**0.056m**

2.6 추가 실험 : 저텍스처 환경 성능 평가

Table 3는 텍스처가 부족하거나 조명이 약한 환경에 대한 데이터셋에서 실험한 결과이다. 기존 ORB-SLAM2[1] 대비 절대 궤적 오차(ATE)가

90% 이상 감소하였으며, 이는 제안 구조가 실환경 적용에서도 높은 견고성과 정밀도를 유지함을 입증한다.

Ⅲ. 결론

본 연구에서는 엣지-클라우드 협업 기반의 분산형 단안 VSLAM 아키텍처를 제안하고, 이에 따른 기술적 기여를 제시하였다. 먼저, ORB-SLAM2[1]와 Depth Anything v2[2]를 각각 엣지와 클라우드에 분산 배치함으로써, 저전력 환경에서도 정밀하고 실시간성이 뛰어난 VSLAM 구조를 구현하였다. 이를 통해 실시간 분산형 VSLAM의 효과적인 구조를 실현할 수 있었으며, 엣지 단에서는 연산 부담을 줄이고, 클라우드에서는 정밀한 깊이 추정 처리를 수행함으로써 각 구성 요소의 장점을 극대화하였다. 또한, 통신 및 연산 최적화를 위해 전체 680프레임 중 초기화 프레임 8개와 키프레임 57개만을 선택적으로 전송하는 방식을 도입하였다. 이 전략을 통해 약 90% 이상의 통신 데이터량을 절감할 수 있었으며, 엣지 디바이스 상에서도 평균 18.8 FPS 이상의 성능을 유지하여 실시간 운용 기준을 만족하였다. 이를 통해 제한된 자원을 가진 환경에서도 안정적이고 효율적인 SLAM 수행이 가능함을 실험적으로 입증하였다. 정밀한 위치 추정 성능 확보 측면에서도 의미 있는 성과를 보였다. TUM RGB-D 데이터셋[3]을 활용한 실험 결과, 기존 ORB-SLAM2[1] 대비 절대 궤적 오차(ATE)가 0.442m에서 0.024m로 약 94.6% 감소하였으며, 조도가 낮거나 텍스처 정보가 부족한 환경에서도 안정적인 동작을 유지하였다. 이는 제안된 구조가 다양한 실내외 환경에서의 정밀한 위치 추정 요구를 충족시킬 수 있음을 보여준다. 종합적으로, 본 논문에서 제안한 분산형 VSLAM 구조는 GPU가 탑재되지 않은 경량 로봇 환경에서도 실질적인 적용이 가능하며, 자율주행 시스템, 모바일 로봇, 실내 내비게이션 등 다양한 응용 분야에서 실용적인 솔루션으로 활용될 수 있다.

향후 연구 방향으로는 세 가지 확장을 고려할 수 있다. 첫째, IMU나 LiDAR와 같은 다양한 센서를 융합한 멀티 센서 기반 구조로의 확장을 통해 더욱 견고하고 정밀한 위치 추정이 가능하도록 할 수 있다. 둘째, 복수의 로봇이 동시에 클라우드와 통신하는 분산 환경을 고려하여 서버 자원의 최적화 및 병렬 처리 구조에 대한 연구가 필요하다. 셋째, 네트워크 상황에 따라 프레임 전송 정책을 동적으로 조정할 수 있는 적응형 통신 알고리즘을 개발함으로써, 다양한 통신 환경에서도 안정적인 실시간성을 확보할 수 있다. 이러한 후속 연구는 본 논문에서 제안한 분산형 VSLAM 구조를 보다 다양한 실제 환경에 효과적으로 확장하고 적용할 수 있는 기반이 될 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.RS-2023-00230661, 하이브리드 양자키분배 방법 및 망 관리 기술 표준개발)과 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원 (P0024177, 2023년 지역혁신클러스터육성)을 받아 수행된 연구임

참 고 문 헌

[1]MUR-ARTAL, Raul; TARDÓS, Juan D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE transactions on robotics, 2017, 33.5: 1255-1262.

[2]YANG, Lihe, et al. Depth anything v2. Advances in Neural Information Processing Systems, 2025, 37: 21875-21911.

[3]STEINBRÜCKER, Frank; STURM, Jürgen; CREMERS, Daniel. Real-time visual odometry from dense RGB-D images. In: 2011 IEEE international conference on computer vision workshops (ICCV Workshops). IEEE, 2011. p. 719-722.