

LLM 기반 적응형 의사결정 로봇 시스템 설계

박효준, 박창균, 김예찬, 정서현, 이지석, 이우석, 임현우, 임진섭, 주민철
국민대학교

{joon_1234, pchb00, kimyechan307, sh040330, runseok23, wolfman352, dla418, limsk519,
mcju}@kookmin.ac.kr

LLM-Based Adaptive Decision-Making Robot System Design

Park Hyojun, Park Changkyun, Kim Yechan, Jeong Seohyeon, Lee Jiseok, Lee Wooseok,
Lim Hyeonwoo, Lim Jinseop, and Ju MinChul
Kookmin Univ.

요약

본 논문은 자연어 명령을 이해하고 물체를 인식해 파지하는 자율 모바일 로봇 시스템을 제안한다. LLM 기반 명령 해석, 딥러닝 기반 시각 인지, 좌표 추정 및 로봇 제어를 통합 프레임워크로 구현하였다. 실험을 통해 다양한 환경에서 실시간 동작과 높은 정확도를 검증하며 서비스 로봇의 실용 가능성을 확인하였다.

I. 서론

최근 대규모 언어 모델(LLM)과 로봇 기술의 융합은 직관적인 사용자 인터페이스와 높은 자동화 가능성으로 주목받고 있다 [1]. 그 중, 사용자의 자연어 지시를 이해하여 움직이고 물체를 집는 모바일 매니퓰레이터는 서비스 로봇 분야의 핵심 과제이다. 이를 위해서는 언어 해석, 3 차원 시각 인지, 이동, 파지 제어가 실시간으로 연결된 신뢰성 높은 구조가 필요하다 [2], [3]. 본 연구는 이러한 기술적 흐름 속에서, 사용자의 자연어 명령을 기반으로 복잡한 인식 및 조작 과정을 자동으로 수행할 수 있는 지능형 로봇 시스템의 구현 가능성을 탐색한다.

II. 제안하는 시스템

본 논문에서는 LLM 과 깊이 카메라를 결합하여 객체 좌표를 추정하고, 이를 기반으로 로봇의 이동과 부착된 로봇 팔의 파지(grasp)를 통합적으로 제어하는 방법을 제시한다. Fine-Tuned LLM 을 통해 사용자의 명령으로부터 로봇의 이동 정보와 객체의 정보를 추출하여 로봇으로 전달한다. TurtleBot3 Waffle 로봇(이하, Waffle)이 요구된 장소로의 Navigation 을 수행한다. 이후, 로봇 팔 상단에 장착된 카메라를 통해 이미지 상의 물체를 인지하는 동작을 수행한다. 이때, 물체가 인지되면 물체의 깊이 정보와 카메라 내부 파라미터를 이용해 카메라 좌표계 기준의 물체의 3 차원 좌표를 추출한다. 마지막으로, Waffle 의 로봇 좌표계 및 월드 좌표계로의 변환 과정을 통해 로봇 이동 혹은 로봇 팔 제어를 수행한다.

2.1 LLM 기반 로봇 자율주행

본 논문에서는 로봇 제어 명령 생성을 위해 LLM 을 활용하였다. 초기에는 OpenAI API 기반의 명령 생성 방식을 사용하였으나, 안정성 확보 및 네트워크 종속성 제거를 위해 로컬 환경에서 직접 파인튜닝한 LLM 으로

대체하였다. 이를 위해 Huggingface Transformers 기반 모델을 외부 장치에서 local로 파인 튜닝하여 운용하였고, 사용자로부터의 텍스트 명령을 처리한 후 ROS 통신을 통해 토픽으로 Waffle 에게 전달되도록 구성하였다.

로컬 학습은 커스텀 명령어-행동 매핑 데이터셋을 기반으로 수행하였으며, 그림 1 은 학습 중 손실 함수(Loss)의 감소 추이를 나타낸다.

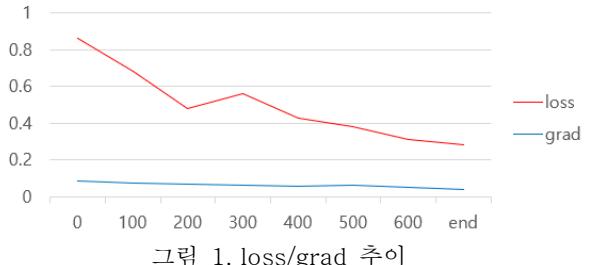


그림 1. loss/grad 추이

이러한 LLM 출력을 활용해, Waffle 은 지정된 목적지까지의 자율주행을 수행하며, 이후 로봇 팔 제어 및 객체 파지 동작으로 연계된다.

2.2 카메라를 이용한 객체 인지 및 좌표 추출

본 논문에서는 객체 인지를 위해 CNN 기반의 딥러닝 모델을 사용하였다. 이를 통해 감지된 물체 중 목표 물체를 식별하여 선택한다. RGB 영상에서 검출된 물체의 중심 픽셀 좌표 (x_p, y_p) 을 기준으로, 해당 위치 주변의 깊이 값을 카메라로부터 수집한다. 이때, 노이즈 완화를 위해 주변의 깊이 값의 평균을 계산하였다. 이후, 추출된 깊이 값 Z 와 카메라 내부 파라미터 행렬 K 를 수식 1에 대입하여 Camera 좌표계 기준 3 차원 좌표 (X_c, Y_c, Z_c) 을 추정하였다. 마지막으로 Camera 좌표계와 World 좌표계의 회전 및 병진 관계를 로봇 플랫폼으로부터 얻어 Robot 좌표계 및 World 좌표계로의 변환을 실시한다. 이렇게 얻어진 좌표 값은 로봇의 이동 혹은 로봇 팔 제어에 활용된다.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = Z \cdot K^{-1} \cdot \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}$$

2.3 로봇팔 제어 알고리즘

로봇팔은 4-DOF로 구성된 OpenMANIPULATOR-X를 사용하였다. Moveit2에 포함된 MotionPlanning, IK, Trajectory Execution 알고리즘을 사용하여 경로 생성, joint 좌표 계산, PID를 제어하였다. 또한 개조한 Waffle의 무게중심과 joint 최대 회전각을 고려하여 Planning Scene을 수정함으로써 충돌 감지 및 회피 범위를 고려하였다.

III. 구현 환경 및 결과

본 논문에서는 TurtleBot3 Waffle에 장착된 Jetson AGX Orin 보드를 통해 딥러닝 기반 실시간 객체 인식 및 로봇 팔 제어를 수행하였다. 자율주행 경로 생성에 필요한 명령은 외부에서 실행 중인 LLM에서 생성되며, 이는 5G 기반 O-RAN 통신을 통해 Jetson 보드에 전달되었다.

3.1 LLM Fine-tuning

본 연구에서는 외부 컴퓨팅 장비에 구축된 로컬 LLM을 통해 사용자 명령어를 처리하였다. 사용자는 그림 2와 같이 자연어 형태로 명령을 입력하며, LLM은 이를 구조화된 그림 3과 같은 JSON 포맷으로 변환한다. 변환된 JSON 명령은 ROS 메시지 형태로 변환된 후, 토픽을 통해 Waffle에 전달된다.

```
Enter a text command: Go backward for 4.1 meters with a speed of 0.3 meters per second.
```

그림 2. 명령어 입력

```
Enter a text command: Go backward for 4.1 meters with a speed of 0.3 meters per second.
[INFO] [1747320183.962794608] [rospt_client]: Response: {"action": "move", "params": {"linear_speed": 0.3, "distance": 4.1, "is_forward": false, "unit": "meter"}}
[INFO] [1747320183.963307218] [rospt_client]: JSON: {"action": "move", "params": {"linear_speed": 0.3, "distance": 4.1, "is_forward": false, "unit": "meter"}}
```

그림 3. JSON 형식의 출력 데이터

3.2 물체 인지 및 3차원 좌표 변환

본 논문에서는 Intel 사의 D435i 카메라를 사용하여 깊이 이미지와 색상 이미지 정보를 얻었다. 객체 인지를 위한 딥러닝 모델로는 YOLOv8을 사용하였다. 2.2 절에서 기술한 기법을 적용해 카메라 좌표계 기준 물체의 3차원 좌표를 추정하였다. 이때 필요한 각 파라미터는 카메라 및 로봇 플랫폼에서 제공해주는 값을 이용하였다. 최종적으로 좌표를 시각화한 결과는 그림 4와 같다.

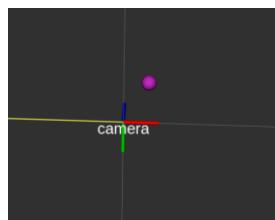


그림 4. 카메라 좌표계 기준 물체의 좌표 시각화

또한 실시간으로 물체의 움직임을 정확하게 3차원 좌표로 추출할 수 있음을 실험적으로 확인하였다. 이를 통해, 로봇이 지속적으로 움직이는 실제 환경에서도 본 논문에서 구현한 시스템을 적용할 수 있음을 확인하였다.

3.3 좌표 기반 로봇팔 모션 플래닝 및 제어

앞선 3.2 절에서 추정한 3차원 목표 좌표를 기반으로 로봇 팔의 목표 위치를 설정한다. MoveIt 2는 입력된 좌표에 대해 역기구학 계산을 수행하여 관절 각도를

결정하고, MotionPlanning 알고리즘으로 충돌 없는 경로를 계획한다. 이후 JointTrajectory 메시지를 생성하여 OpenCR 보드에 전달하고, 로봇 팔을 제어한다. 목표 위치에 도달하면 그리퍼를 작동시켜 물체를 파지하고, 다시 초기 위치로 자동 복귀함으로써 전체 작업 사이클을 완성한다.

그림 5는 목표 물체를 추적하여 로봇 팔 제어를 통해 물체를 파지한 상황에 대한 사진이다.



그림 5. 로봇 팔 제어를 통한 물체 파지

이와 같은 과정을 통해, LLM으로 자연어 명령을 해석하여 로봇 이동, 목표 물체를 인식 후 파지하는 과정이 하나의 프레임워크로 진행되었다.

IV. 결론

위 연구는 언어 이해 기반의 명령 처리부터 객체 인식, 3차원 좌표 추정, 로봇의 이동 및 팔 제어에 이르기까지, 일련의 과정을 하나의 통합된 프레임워크로 실현한 점에서 큰 의의를 가진다. 각 요소 기술을 단순히 나열한 것이 아니라, 모듈 간 연동성과 실시간성을 확보하여 실제 작동 가능한 시스템으로 구현했으며, 다양한 환경에서도 안정적이고 정밀한 동작을 확인하였다. 이를 통해 로봇이 단순 반복 동작을 수행하는 수준을 넘어, 의미를 해석하고 상황에 맞춰 유연하게 반응하는 방향으로 진화할 수 있음을 보여주었다.

본 논문은 향후 물류 지원, 가정용 서비스, 재난 현장 등 다양한 도메인에서 인간의 부담을 줄이는 지능형 로봇 시스템 구축에 실질적 근거를 제공한다.

ACKNOWLEDGMENT

본 과제(결과물)는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 첨단분야의 혁신융합대학사업(차세대통신)의 연구 결과입니다.

참고 문헌

- [1] 이준기, 박상준, 김나우, 김에덴, 고석갑, "거대언어모델 기반 로봇 인공지능 기술 동향," ETRI Trends, 2024년 2월, pp. 95-105.
- [2] L. Kunze et al., "Artificial intelligence for long-term robot autonomy: A survey," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, 2018, pp. 4023-4030.
- [3] M. Brady, "Artificial intelligence and robotics," *Artif. Intell.*, vol. 26, no. 1, 1985, pp. 79-121.