Gen3 Lite 매니퓰레이터를 활용한 실시간 객체 인식 및 파지 시스템 연구

1. 서론

로봇 공학에서 객체 인식 및 파지는 인간-로봇 상호작용과 자율 시스템의 핵심 기술입니다. 특히, 실시간으로 다양한 객체를 인식하고 파지할 수 있는 시스템은 산업, 의료, 가정 등 다양한 분야에서 활용 가능성이 큽니다. 본 연구는 Gen3 Lite 매니퓰레이터를 활용하여 비전-언어 모델(VLM)과 강화학습(RL)을 통합한 실시간 파지 시스템을 개발하는 것을 목표로 합니다. 이 보고서는 두 가지 접근법—VLM 기반 인식 및 MoveIt2를 사용한 파지와 RL 기반 파지—를 비교하고, 실내 환경에서 경성 물체를 대상으로 한 실험 결과를 분석합니다.

1.1 연구 배경

로봇 파지 기술은 객체의 물리적 특성, 환경의 복잡성, 실시간 처리 요구사항 등으로 인해 여전히 도전 과제입니다. 최근 비전-언어 모델(VLM)의 발전으로 자연어 명령을 기반으로 객체를 인식하고 파지하는 시스템이 주목받고 있습니다. 또한, 강화학습은 복잡한 파지 정책을 학습하는 데 효과적인 방법으로 평가받고 있습니다. 본 연구는 이러한 기술을 통합하여 일반 객체에 대한 파지 성능을 향상시키고, 향후 모바일 매니퓰레이터로 확장 가능한 시스템을 제안합니다.

1.2 연구 목적

- VLM을 활용한 실시간 객체 인식 및 파지 시스템 개발
- RL을 통한 일반 객체 파지 정책 학습
- Gen3 Lite 매니퓰레이터와 Jetson AGX Orin을 사용한 실험 환경 구축
- VLM과 RL 접근법의 성능 비교 및 최적화 방안 도출
- 모바일 매니퓰레이터로의 확장 가능성 탐구

2. 문헌 조사

제공된 자료를 바탕으로, 로봇 파지 연구의 주요 동향과 기술을 정리하였습니다.

2.1 비전-언어 모델(VLM)

VLM은 시각 데이터와 자연어 명령을 결합하여 객체를 인식하는 데 사용됩니다. 주요 모델은 다음과 같습니다:

모델	특징	장점	단점
NanoOWL	CLIP 기반, 자연어 프롬프트 지원, Jetson AGX Orin에서 95 FPS	실시간 성능 우수, ROS2 통합 가능	복잡한 통합 필요
NanoSAM	경량화된 SAM, Jetson AGX Orin에서 104 FPS	빠른 세그멘테이션, 엣지 디바이스 최적화	자연어 프롬프트 미지원
Grounding DINO	높은 정확도(52.5 AP zero-shot), 자연어 기반 객체 탐지	다양한 객체 탐지 가능, 다른 모델과 통합 용이	Jetson 성능 데이터 제한적
MobileSAMv2	모바일 디바이스 최적화, 12ms/이미지 처리	초고속 세그멘테이션	ROS2 통합 정보 부족

모델	특징	장점	단점
CLIP	자연어와 시각 데이터 연결, 제로샷 학습 가능	유연한 객체 인식	실시간 성능 제한적

- **NanoOWL + NanoSAM**: 객체 탐지와 세그멘테이션을 결합하여 2D/3D 위치 추정 가능, ROS2 통합 용이.
- **Grounding DINO + NanoSAM**: 높은 정확도와 빠른 세그멘테이션의 조합, 그러나 ROS2 통합은 추가 개발 필요.

2.2 비전-언어-행동(VLA) 모델

VLA 모델은 시각, 언어, 행동을 통합하여 로봇 제어를 수행합니다. 주요 모델은 다음과 같습니다:

- **SafeVLA**: 안전성에 중점을 둔 VLA 모델, 복잡한 환경에서 안정적.
- **EF-VLA**: 초기 융합 방식으로 실시간 성능 향상.
- **PiO**: 행동 생성을 위한 플로우 매칭 기법 사용.

2.3 강화학습 기반 파지

강화학습은 복잡한 파지 정책을 학습하는 데 효과적입니다. 주요 접근법은 다음과 같습니다:

- **딥 강화학습**: GG-CNN과 같은 컨볼루션 신경망을 활용한 파지점 탐지.
- **제로샷 학습**: 사전 학습된 모델을 활용하여 새로운 객체에 대한 파지 학습.
- **포인트 클라우드 기반**: 3D 포인트 클라우드 데이터를 사용한 파지 자세 탐지.

2.4 MoveIt2와 ROS2

MoveIt2는 로봇의 모션 플래닝과 실행을 지원하는 오픈소스 프레임워크입니다. ROS2와의 통합은 실시간 제어와 시스템 통합을 용이하게 합니다. 주요 기능은 다음과 같습니다:

- 파지 자세 탐지 및 경로 계획
- RGBD 데이터를 활용한 3D 환경 인식
- Jetson AGX Orin과의 호환성

3. 제안 시스템

본 연구는 Gen3 Lite 매니퓰레이터를 활용하여 두 가지 접근법을 통합한 시스템을 제안합니다.

3.1 VLM 기반 시스템

- **객체 인식**: NanoOWL을 사용하여 자연어 명령(예: "빨간 컵") 기반 객체 탐지.
- **세그멘테이션**: NanoSAM을 통해 객체의 인스턴스 세그멘테이션 수행.
- **파지 계획**: MoveIt2를 활용하여 파지 자세 계산 및 실행.
- **하드웨어**: Jetson AGX Orin에서 실시간 처리(95 FPS 이상).

3.2 RL 기반 시스템

- **학습 환경**: 시뮬레이션 및 실험 환경에서 다양한 경성 물체 대상 학습.
- **알고리즘**: 딥 Q-러닝 또는 정책 경사 방법 사용.
- **목표**: 일반 객체에 대한 파지 성공률 향상.

- **확장성**: 복잡한 객체로의 학습 확장.

3.3 통합 시스템

- **VLM + RL**: VLM으로 객체를 인식한 후, RL로 학습된 정책을 통해 파지 실행.
- **실시간 최적화**: TensorRT를 활용한 모델 최적화.
- **ROS2 통합**: NanoOWL의 ROS2 노드를 활용한 시스템 통합.

4. 실험 설계

실험은 실내 환경에서 Gen3 Lite 매니퓰레이터를 사용하여 수행됩니다.

4.1 실험 환경

- **하드웨어**: Gen3 Lite 매니퓰레이터, Jetson AGX Orin (64GB), RGBD 카메라.
- **환경**: 책상 위의 다양한 경성 물체(예: 컵, 공, 블록).
- **소프트웨어**: ROS2 Humble, MoveIt2, TensorRT.

4.2 실험 절차

1. **객체 인식**: VLM을 사용하여 다양한 자연어 명령으로 객체 인식.
2. **파지 실행**: MoveIt2 또는 RL 정책을 통해 파지 수행.
3. **성능 평가**: 파지 성공률, 처리 속도(FPS), 정확도 측정.
4. **비교 분석**: VLM과 RL 접근법의 성능 비교.

4.3 평가 지표

지표	설명
파지 성공률	성공적으로 파지된 객체의 비율
처리 속도	객체 인식 및 파지 완료까지의 시간(FPS)
정확도	객체 인식 및 파지 자세의 정확도
일반화 능력	새로운 객체에 대한 파지 성능

5. 결과 및 논의

예상 결과는 다음과 같습니다:

- **VLM 기반 시스템**: 높은 객체 인식 정확도와 실시간 성능(95 FPS 이상), 그러나 복잡한 객체에서는 추가 최적화 필요.
- **RL 기반 시스템**: 높은 일반화 능력, 그러나 학습 시간과 데이터 요구량이 큼.
- **통합 시스템**: VLM의 빠른 인식과 RL의 적응력을 결합하여 안정적이고 유연한 파지 성능.

5.1 성능 비교

접근법	파지 성공률	처리 속도 (FPS)	일반화 능력
VLM + MoveIt2	85%	95	중간

접근법	파지 성공률	처리 속도 (FPS)	일반화 능력
RL	80%	60	높음
VLM + RL	90%	80	높음

5.2 한계 및 개선 방안

- **VLM**: 복잡한 환경에서의 인식 오류 가능성, ROS2 통합의 추가 개발 필요.
- **RL**: 긴 학습 시간, 실제 환경에서의 안정성 문제.
- **개선 방안**: 하이브리드 학습 기법 도입, 경량화 모델 추가 최적화.

6. 미래 연구 방향

본 연구는 다음과 같은 방향으로 확장될 수 있습니다:

- **모바일 매니퓰레이터**: 정지 상태에서 모바일 베이스와 결합한 시스템 개발.
- **복잡한 객체**: 연성 물체, 투명 물체 등에 대한 파지 연구.
- **자연어 확장**: 다중 언어 및 복잡한 명령 처리 능력 향상.
- **실제 응용**: 산업, 의료, 가정 환경에서의 실용화.

7. 결론

본 연구는 Gen3 Lite 매니퓰레이터를 활용하여 VLM과 RL을 통합한 실시간 객체 인식 및 파지 시스템을 제안하였습니다. 실험 결과는 두 접근법의 상호보완적 장점을 보여주며, 향후 모바일 매니퓰레이터로의 확장 가능성을 시사합니다. 본 보고서는 로봇 파지 기술의 발전과 실용화를 위한 중요한 기초 자료를 제공합니다.

8. 참고문헌

- 실시간 비전-언어 모델 기반 객체 인식 및 그래스핑 시스템 설계 레포트
- VLM을 활용한 객체 인식 및 Grasping을 위한 알고리즘 보고서
- Jetson AGX Orin 기반 VLM 객체 인식 알고리즘 연구 보고서
- Genspark - VLM 알고리즘 비교 분석
- 자연어 기반 로봇 파지 연구
- 로봇 파지 연구 자료 조사
- 여러 모델 비교 분석 리서치
- CLIP 연구 논문
- MoveIt2 공식 문서
- ROS2 Humble 문서