앱 카메라 기반 손동작 인식과 로봇 제어 시스템

최진, 정진교, 김영동*, 현장훈*

국립한밭대학교

20221080@edu.hanbat.ac.kr, 20227004@edu.hanbat.ac.kr, *ydkim1293@hanbat.ac.kr,

*jhhyeon@hanbat.ac.kr

App camera-based hand motion recognition and robot control system

Jin Choi, Jinkyo Jeong, Youngdong Kim*, Janghun Hyeon*

Hanbat National Univ., *Hanbat National Univ.

요 약

본 논문은 터치·조이스틱 중심의 기존 로봇 제어 방식이 가진 조작 피로도 문제와 착용형 센서 기반 모션 인식의 불편함을 개선하기 위해, 모바일 앱 카메라 기반 손동작 인식 로봇 제어 시스템을 제안한다. Fine-Tuning 한 YOLOv11n 모델을 통해 정지·전진·후진·특수 동작을 실시간으로 분류하고, Bounding Box의 위치와 크기를 활용해 로봇의 회전 및 속도를 직관적으로 제어한다. 또한 Flutter 기반 애플리케이션을 통해 카메라 스트리밍, 제스처 인식 결과 시각화 기능을 제공하며 제스처 조합을 이용한 커맨드 기반 행동 제어를 통해 다양한 동작 수행이 가능하도록 하였다. 실험 결과 설계한 시스템은 실시간 환경에서 약 50fps로 안정적인 동작이 가능했으며 제스처 인식 정확도는 mAP50에서 0.995로 높은 정확도와 안정성을 보였다.

Ⅰ. 서 론

본 논문에서는 터치 입력, 조이스틱, 버튼 기반 제어의 조작 피로도 문제와 관성 센서(IMU) 등을 활용한 웨어러블 방식이 갖는 장착 과정의 번거로움과 착용 불편함 문제, 손가락 관절 기반 인식 기법의 높은 연산 비용과 배경 변화의 취약성 문제를 해결하기 위해 애플리케이션 카메라를 이용한 손동작 인식 기반 로봇 제어 시스템을 제안한다.[1][2][3][4] YOLOv11n 모델을 사용자가 직접 구축한 소규모 데이터셋으로 Fine-Tuning하여 정지·전진·후진·특수 동작을 실시간으로 분류하고, Bounding Box의 위치와 크기를 활용하여 회전 및 속도를 제어할 수 있도록 설계하였다. 또한 단일 제스처 분류뿐만 아니라 여러 제스처 조합을 이용하여 더욱 다양한 행동을 수행하는 커맨드 조합 기반 제어 방식을 설계하였다. 마지막으로 Flutter 기반의 앱에서 실시간 카메라 스트리밍과 인식 결과를 제공하여, 로봇의 상태와 명령을 쉽게 확인하고 별도의 장비 없이 제어할 수 있다.

Ⅱ. 본론

2.1 시스템 개요

본 논문에서 설계한 손동작 기반 로봇 제어 시스템은 모바일 애플리케이션 카메라로부터 입력된 영상을 실시간으로 처리하여 제스처를 인식하고, 제스처를 로봇 동작으로 변환하는 하나의 파이프라인으로 구성되어 있다. 시스템은 영상 획득, 제스처 인식, 제어 로직, 로봇 제어, 사용자 인터페이스로 구성하였으며 각 모듈은 독립적으로 동작하도록 설계하였다. 모바일 앱에서 촬영한 영상은 YOLOv11n 기반 제스처 인식 모듈에서 네 가지 동작(정지·전진·후진·특수 동작)으로 구분되며 Bounding box의 위치와 크기 정보를 기반으로 회전 및 속도를 제어한다. 그 후 로봇 제어 모듈에서 명령을 생성하고, 생성된 명령은 ROS2 메시지를 통해 Turtlebot3에 전달되어 동작을 수행한다.[5] 전체 구조도는 그림 1과 같다.

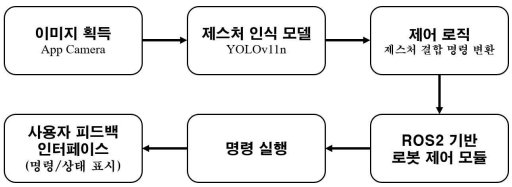

그림1. 전체 파이프라인 구조도

2.2 제스처 인식 모델

제스처 인식 모듈은 YOLOv11n 기반 객체 검출 모델을 기반으로 설계되었으며, 모바일 환경에서의 실시간 처리를 목표로 하였기에 경량화된 구조를 사용하였다.[6] 모델은 연구에서 제작한 forward 500장, backward 220장, special 220장, stop 340장의 이미지, 총 1,280장의 제스처 데이터셋 학습을 통해 네 가지 손동작(정지, 전진, 후진, 특수 동작)을 구분하고, 영상 입력으로부터 손 위치(Bounding Box)와 제스처 클래스를 검출한다.

실험 결과 YOLOv11n 모델을 Fine-Tuning 하였을 때 mAP50 기준 약 0.99의 정확도를 보였다. 이를 통해 정지·전진·후진·특수 동작의 각 클래스는 혼동 없이 안정적으로 구분될 수 있음을 확인하였다. 자세한 실험 결과는 표 1과 그림 2와 같다.

손 위치(Bounding Box) 추정 역시 mAP50-95에서 0.9 이상의 성능을 보여 후속 제어 단계에서 회전 및 속도 계산에 필요한 좌표 정보를 안정적으로 제공할 수 있음을 확인하였다.

표1. 제스처 인식 성능

Class	Images	Instances	Precision (P)	Recall (R)	mAP@0.5	mAP@ 0.5:0.95
All	170	170	0.961	0.961	0.995	0.902
Forward	50	50	0.992	1.000	0.995	0.896
Backward	40	40	1.000	0.845	0.995	0.928
Stop	40	40	0.985	1.000	0.995	0.869
Special	40	40	0.868	1.000	0.995	0.914

그림2. 제스처 인식 test class 판별 결과

2.3 로봇 제어 로직

제어 로직은 제스처 인식 결과와 손의 위치 정보를 기반으로 로봇의 이동 방향과 속도를 결정한다. 먼저 제스처 인식 모듈에서 검출된 세 가지 동작(정지, 전진, 후진)은 직선 이동 명령을 지정하고, 정지 동작은 모든 속도를 0으로 하여 즉시 정지하도록 한다. 또한 실시간 추론 중 일부 프레임에서 발생하는 오탐지로 인한 제어 불안정 문제를 해결하기 위해, 동일 제스처 클래스가 일정 시간(0.5 초) 이상 연속으로 유지되는 경우에만 해당 명령을 입력으로 하도록 설계하였다.

Bounding Box의 좌표는 화면 중심을 기준으로 좌우 위치 편차를 계산하여 손이 화면의 왼쪽에 위치하면 좌회전, 오른쪽에 위치하면 우회전 방향으로 회전 속도를 더한다. 또한 Bounding Box의 크기는 손과 카메라 간의 거리로 판단하여, 전진·후진 시 이동속도를 조절하는 데 사용된다.

추가로 본 연구에서는 단일 제스처 분류를 기반으로 한 단순 제어 방식에서 확장하여, 제스처 지속 시간 및 조합을 고려하는 커맨드 기반 제어를 설계하였다. 예를 들어 특수 동작 제스처가 일정 시간 이상 유지된 상태에서 전진 제스처가 감지되었을 때 제자리 회전 후 전진과 같은 추가적인 동작을 수행할 수 있도록 하였다. 이런 방식은 제한된 제스처 개수로도 다양한 행동을 구현할 수 있다는 장점이 있다.

2.4 모바일 애플리케이션 및 통신 구조

본 연구에서 모바일 애플리케이션은 제스처 기반 로봇 제어 시스템의 입력장치이자 사용자 인터페이스 역할을 수행한다. Flutter를 기반으로 설계한 애플리케이션은 스마트폰의 카메라로부터 실시간 영상을 획득하고 획득한 영상을 제스처 인식 모듈로 전송하여 그림 3과 같이 인식 결과를 사용자에게 시각적으로 제공하도록 설계되었다.

영상 전송과 제어 명령 전송은 WiFi 기반의 공유된 네트워크를 통해 이루어졌으며, rosbridge websocket을 활용하여 모바일 애플리케이션과 ROS2 시스템 간 실시간 데이터를 송수신한다.

영상 전송, 데이터 인식 결과 전달, 제어 명령 전송 과정에서 지연으로 인한 불안정한 모습은 관찰되지 않았으며, 전체 시스템은 약 50fps 수준의 실시간 성능을 유지하였다. 이를 통해 제안한 모바일-로봇 연동 구조가 실시간 제스처 기반 제어에 충분한 통신 성능을 제공함을 확인하였다.

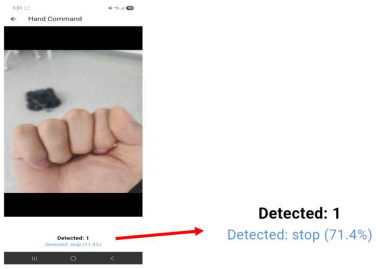

그림3. 애플리케이션 class 탐지 결과

Ⅲ. 결론 및 향후 연구

본 논문에서는 모바일 애플리케이션 카메라 기반 손동작 인식과 로봇 제어를 결합한 손동작 인식 기반 로봇 제어 시스템을 제안하였다. YOLOv11n 모델을 소규모 데이터셋으로 Fine-Tuning 하였을 때 정지, 전진, 후진, 특수 동작, 총 네 가지 손동작을 높은 정확도로 인식함을 확인하였으며, Bounding Box 위치 및 크기 정보를 활용한 로봇 제어를 통해 직관적인 로봇 제어가 가능하였다. 또한 제스처 조합을 고려한 커맨드 기반 제어 방법을 설계함으로써 단순한 제스처 분류를 넘어 더욱 다양한 동작 수행이 가능함을 확인하였다.

향후 연구에는 단순한 커맨드 설계가 아닌 제스처, 이동 패턴, 시간적 변화를 함께 고려하는 고급 커맨드 설계로 확장할 계획이다. 또한 복잡한 배경이나 조명 변화에 대한 강인성을 확보하기 위해 손 영역 분할 기반의 배경 제거 기법을 적용한 제스처 인식 모델로 확장할 계획이다.

참 고 문 헌

[1] J.Shotton et al., “Skeleton Tracking Method and Apparatus,” U.S. Patent US8254634B2, Microsoft Corporation, Aug. 21, 2012.

[2] KHAN, Naimat Ullah; WAN, Wanggen. A review of human pose estimation from single image. In: 2018 international conference on audio, language and image processing (ICALIP). IEEE, 2018. p. 230-236.

[3] LIANG, Ci-Jyun, et al. A vision-based marker-less pose estimation system for articulated construction robots. Automation in Construction, 2019, 104: 80-94.

[4] ZHANG, Fan, et al. Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214, 2020.

[5] MACENSKI, Steve, et al. From the desks of ROS maintainers: A survey of modern & capable mobile robotics algorithms in the robot operating system 2. Robotics and Autonomous Systems, 2023, 168: 104493.

[6] KHANAM, Rahima; HUSSAIN, Muhammad. Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725, 2024.