

딥 뉴럴 네트워크 및 손 추적 기반의 웨어리스 IoT 장치 컨트롤러

최승준*, 김은열*, 김정화*, 황채은*, 최태영†

*금오공과대학교 컴퓨터공학과

e-mail: easdkr@kumoh.ac.kr

Wearless IoT Device Controller based on Deep Neural Network and Hand Tracking

Seung-June Choi*, Eun-Yeol Kim*, Jung-Hwa Kim*, Chae-Eun Hwang*,
Tae-Young Choi†

*Dept of Computer Engineering, Kumoh National Institute of Technology

요 약

본 논문에서는 거동이 불편한 환자나 장애인들을 위해 신체에 착용하는 부가적인 장비 없이 멀리 있는 가전을 직접 움직이지 않고 편리하게 제어할 수 있는 RGB-D 카메라를 활용한 손 인식과 딥러닝 기반 IoT 장치 컨트롤 시스템을 제안한다. 특히, 제어하고자 하는 장치의 위치를 알기 위하여 YOLO 알고리즘을 이용하여 장치를 인식한다. 또한 그와 동시에 RGB-D 카메라의 라이브러리를 이용하여 사용자의 손을 인식, 현재 사용자 손의 위치와 사용자가 취하는 손동작을 통하여 해당 위치의 장치를 제어한다.

1. 서론

현재 IoT, AI, 빅데이터 등 IT 기술의 발전으로 보다 윤택하고 편리한 생활을 하고 있다. 그러나 이러한 기술들의 비약적인 발전에도 불구하고 거동이 불편한 고령자나 투병인 혹은 장애인들에게는 기술의 혜택이 닿지 않고 있는 것이 현실이다. 그렇기에 거동이 불편한 사람들에게 직접 움직이지 않고도 멀리 있는 가전을 간편하게 조작할 수 있는 보조기기가 있다면 기기를 사용하는 당사자뿐만 아니라 거동을 보조하는 보호자 혹은 간병인에게 상당한 편의를 제공할 수 있으며 의료 복지 효율에 증대를 기대할 수 있다.

본 논문에서는 일상생활에 편의를 줄 수 있는 보조기기를 구현하기 위해 멀리 떨어져 있는 가전(IoT 디바이스)을 손으로 직접 조작하지 않고도 카메라를 통해 작동할 수 있는 시스템을 제안한다. 우선 원거리의 IoT 디바이스들을 후방 카메라로 영상을 입력받아 사용자에게 모니터를 통해 제공한다. 그리고 사용자가 모니터를 통해 인식된 IoT 디바이스를 손동작으로 조작할 수 있는 기능을 제공한다.

사용자는 IoT 디바이스를 조작하기 위해 복잡한 프로그램이나 장비를 조작할 필요 없이 전방의 센서 카메라를 통해 손동작만으로 모든 제어가 가능하며, 이로써 전방과 후방의 영상을 동시에 처리하는 실시간 다중영상처리 시스템을 구현하고 기존의 기능을 합성하여 Hand Remote Control을 실현하는 기술적 특징을 가진다.

2. 관련 연구

IoT 디바이스를 제어하기 위한 연구에는 웨어러블 기기를 이용하는 방법과, 음성인식을 이용하는 방법 등이 있다. 웨어러블 기기를 이용하는 방법에는 터치 센서 및 VR 기기를 이용한 연구 [1]와 손목 시계형 웨어러블 디바이스를 이용하여 사용자의 제스처를 인식하는 방식 [2]이 있다. 두 가지 방법 모두 웨어러블 기기를 이용하기 때문에 몸에 기기를 부착해야 한다는 단점이 있다. 특히 전자의 경우, 증강현실 기기를 사용하여 사용자들이 손목 회전을 통해 기기를 제어하는 방식으로 얼굴의 방향을 조절한 뒤 터치 센서를 조작할 수 있다. 하지만 센서를 누른 상태에서 메뉴 조작 및 초점을 조절해야하는 어려움이 있다.

다른 제어 방법으로 음성인식 기반의 Home Automation 시스템을 구현한 연구가 있다 [3]. 이 연구에서 사용한 음성 인식 시스템은 소음에 따른 환경적 제약이 강하고, 음성 인식기능이 여전히 미흡하여 일상생활에서 사용하기에 다소 불편한 점이 있다.

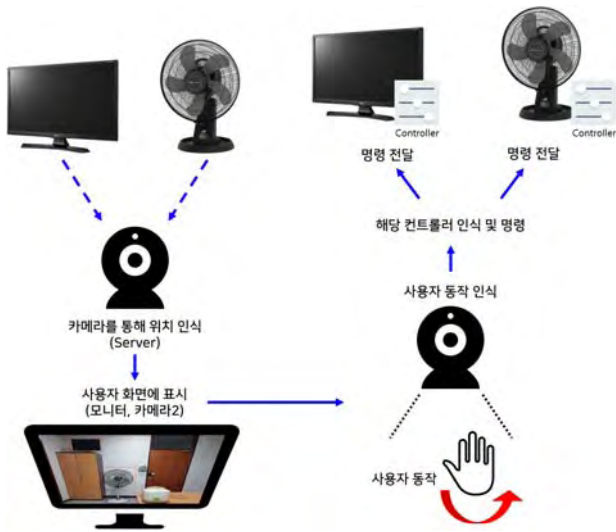
본 연구에서는 이러한 불편한 점들을 개선하기 위해 웨어리스 기반의 IoT 장치 컨트롤러를 제안한다.

* 공동 1저자로서 본 논문에 동일하게 기여하였음.

† 교신저자

3. 제안 시스템

3.1 시스템 구조



(그림 1) 시스템 구조

Hand Gesture로 IoT 디바이스를 제어하는 시스템을 구축하기 위해 본 논문에서는 크게 두 개의 프로세스를 구축했다. RGB-D 카메라를 이용해 영상 내의 손의 좌표와 Gesture Recognition 결과를 서버(Object Detection Process)로 보내는 프로세스와, YOLO 알고리즘을 이용해 Object Detection하여 획득한 영상 내 IoT 디바이스의 좌표와 Hand Tracking 결과를 이용해 IoT 디바이스를 인식하고, 제어하기 위한 메시지를 보내는 프로세스로, 두 프로세스는 서로 소켓 통신하도록 설계하였다.

제안하는 시스템을 통한 IoT 디바이스 제어의 예는 다음과 같다. 실시간으로 실내 환경을 Object Detection한 결과를 출력한 영상 위에, 사용자의 손 움직임을 따라 Hand Pointer가 움직인다. 이는 사용자의 Hand Tracking 정보를 실시간으로 업데이트 하여 영상에 출력하는 방식으로 동작한다. 인식한 물체 영상 위에 Hand Pointer가 있을 때 사용자가 제스처를 취하면, 제어하기 위한 메시지를 해당 IoT 디바이스에게 JSON 형식으로 전송한다.

3.2 실시간 물체 탐지

본 논문에서는 일반 가정의 거실이나 병실 등 실내의 IoT 디바이스들을 탐지하고 사용자에게 실시간으로 물체가 탐지된 영상을 제공하기 위해 딥러닝 기반의 실시간 Object Detection 알고리즘인 YOLO [4]을 사용한다. YOLO 알고리즘은 하나의 신경망(Neural Network)을 전체 이미지에 적용한다. 이 네트워크는 이미지를 $S \times S$ Grid로 분할한다. 각 Grid Cell은 여러 개의 Bounding box와 각 Bounding Box에 대한 Confidence Score를 갖는다. 이 Confidence Score는 Cell에 Object가 존재하는지에 대한 수치이다. 초기에는 객체와 동떨어진 Bounding Box가 그려지고, Confidence Score를 계산하여 가장 객체

인식 정확성이 높은 위치로 Bounding Box의 위치를 조정한다. 또한 각 Grid Cell은 객체 포함 여부를 계산하기 위해 N 개의 Confidential Class Probability를 계산한다. 주변의 Grid와 병합하며 신뢰도를 높이고, 임계값을 설정하여 불필요한 부분을 제거함으로써 물체를 인식한다.

YOLO 알고리즘의 경우 GTX 1080ti 환경에서 약 30fps 정도의 성능을 보인다. 기존의 다른 real-time Object detection algorithm에 비해 처리가 단순하고 속도가 매우 빠르기 때문에 본 논문에서는 YOLO 알고리즘을 사용하여 IoT 디바이스의 이미지를 학습하고, 실시간으로 물체를 탐지하여 사용자에게 제공한다.

3.3 손 위치 추적

본 논문에서 사용자의 손 위치를 추적하기 위해 Intel RGB-D 카메라 SR300을 이용한다. [5] SR300 카메라는 전방 카메라로 사용한다. 전방 카메라로 사용자를 촬영하면서 사용자의 손을 인식한다. 사용자의 손을 인식하여 사용자가 손을 움직일 때마다 전방 카메라 화면에 OpenCV를 이용하여 사용자의 손 좌표를 마우스 포인터로 표시한다. 사용자는 화면에 보이는 마우스 포인터로 자신의 손 위치를 확인할 수 있다.

사용자가 본 연구에서 사용하는 손 제스처는 'Click', 'Clockwise', 'Counterclockwise' 세 가지가 있다. 'Click'은 손을 폈다 쥐었다 하는 제스처로 후방 카메라에 보이는 물체를 선택하는 제스처다. 'Clockwise'는 손을 편 채로 시계 방향으로 돌리는 제스처로 선택된 물체의 상태를 'ON'으로 변경하는 제스처이다. 'Counterclockwise'는 손을 편 채로 반시계 방향으로 돌리는 제스처로 선택된 물체의 상태를 'OFF'로 변경하는 제스처이다.

위의 세 가지 제스처로 사용자가 물건을 선택하거나 선택한 물체의 상태를 변경한다. 이를 통해 사용자는 물체를 직접 만지지 않고 편리하게 제어가 가능하다.

4. 실험 결과 및 분석



(그림 2) 커서가 TV 위에 있는 상태



(그림 3) 커서가 밥솥 위에 있는 상태

본 연구에서는 실내에서 사용하는 실제 기기들을 학습하여 실험을 진행하였다. 학습에 사용된 Training data는 표 1과 같으며, Validation data는 각 클래스의 1/5 만큼의 데이터를 사용하였다.

Class	개수 (개)
청소기	625
에어컨	235
밥솥	774
TV	530
세탁기	406
문	180
전체	2,750

<표 1> Training data 개수

NVIDIA GeForce GTX 1080 Ti를 사용한 환경에서 30fps 이상의 성능을 보였으며, 30,000번의 반복 학습 결과 98%의 MAP를 보였다.

그림 2와 그림 3은 마우스 포인터가 물체 위로 움직였을 때, 하단 바에 인식된 물체의 상태를 표시한 그림이다. 화면 하단 상태 표시 창에 명시된 것처럼 YOLO 알고리즘을 이용하여 인식한 물체의 Label을 보여준다.



(그림 4) Clockwise 제스처 표시 화면

그림 4는 사용자가 선택한 물체의 상태를 'ON'으로 변경하고자 하는 제스처를 취하는 그림이다. 사용자가 이러한 제스처를 취하면 선택된 물체에 해당하는 선택한 IoT 디바이스의 상태가 'ON'으로 변경된다.



(그림 5) 오브젝트 상태 변경

그림 5는 사용자가 그림 4에서 보이는 것처럼 제스처를 취하면 선택된 물체의 변경된 상태를 하단에 보여주는 그림이다. 이와 같이 사용자가 'OFF' 제스처를 취하면 하단에 물체의 Status가 'OFF'로 표시된다.

5. 결론 및 향후 연구

본 논문에서는 RGB-D 카메라(Intel SR300)를 사용하여 사용자의 손의 위치와 제스처를 탐지하고 YOLO 알고리즘을 이용해 실시간으로 IoT 장치를 탐지함으로써 신체에 접촉하는 다른 장치의 착용 없이 손동작만으로 멀리 있는 IoT 디바이스들을 제어하는 시스템을 구현하였다. 향후 연구를 통해 사물의 인식률을 높이고 종류를 다양화함으로써 거동이 불편한 장애인이나 환자뿐만 아니라 일반 가정 내에서도 CCTV를 이용해 사용자들이 편리하게 실내의 IoT 장치를 제어할 수 있도록 개선할 예정이다.

참고문헌

- [1] Kyle Koh, Han Joo Chae, Yuri Choi, Jeongin Hwang, Yi Eun Kim and Jinwook Seo, "A Study on User Interface for Device Configuration with Wrist-Rotation Using AR and Wearables in IoT Environment", 한국정보과학회 학술발표논문집, 2016.6, 1411-1413
- [2] Jun Su Kim, Seung Min Baek, Yeon Ju Jin, Sang Wook Han, and Jin-Woo Jeong, "Design and Implementation of IoT Control System Based on Gestures with Wearable Devices", The Journal of Korean Institute of Information Technology 15, 2017.10, 101-112
- [3] H. Rajput, K. Sawant, D. Shetty, P. Shukla, and A. Chougule, "Implementation of Voice Based Home

- Automation System Using Raspberry Pi,” Int. Res. J. Eng. Technol., 2018.
- [4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” IEEE CVPR, pp. 779-788, 2016
- [5] Intel RealSense, <https://software.intel.com/en-us/realsense>