

모션인식과 음성인식을 활용한 입력장치 구현

박양우, 이우재, 김민섭, 정명진, 강민재, 염상호

*경운대학교 소프트웨어학부,

°경운대학교 항공소프트웨어공학과

e-mail: ywpark@ikw.ac.kr, n13lue@naver.com, seup1026@naver.com, lillil1111@naver.com,
gusdn2017@naver.com, 539436@naver.com,

Implementation of input device using motion and voice recognition

Yang Woo Park, Woo Jae Lee, Kim Min Sup, Myeong Jin Jeong, Min Jae Kang, Sang Ho Yeom

*Dept. of Software, Kyungwoon University,

°Dept. of Aeronautical Software Engineering, Kyungwoon University

● 요약 ●

본 논문에서는 실시간 영상 처리 라이브러리인 OpenCV와 미디어파이프(MediaPipe)를 사용하여 동작 인식 기반 키보드, 마우스 제어 프로그램을 개발하였다. Google의 미디어파이프(MediaPipe)에서 제공하는 손가락 마디 부분의 랜드마크를 인식하며, 실시간 영상 처리로 띄워진 사용자 인터페이스에서 제스처를 통해 키보드 입력과 마우스 제어를 할 수 있으며, Google에서 제공하는 오픈 소스와 결합하여 음성인식을 통한 키보드 입력이 가능하다. 또한, 각 기능끼리 제스처를 통해 기능 변경이 가능하여 다양한 산업 분야에서 원하는 키를 매핑할 수 있기 때문에 활용 가능성이 높으며, 의료 분야에서 감염 예방을 목적으로 사용할 수 있다. 특히 기존의 메타버스에서 사용되는 고가의 센서를 대체하여 비용 절감 부분에서 장점이 있다.

키워드: 모션인식(Motion Recognition), 음성인식(Voice Recognition), 미디어파이프(MediaPipe)

I. Introduction

본 연구에서는 미디어파이프(MediaPipe)를 이용하여 각 손가락 랜드마크를 가져와 모션 인식 및 제어에 활용한다. 구현한 알고리즘을 통해 키보드 입력과 기능 간 변환을 할 수 있도록 하였다. 웹캠과 마이크를 통해 실시간 영상 처리로 키보드, 마우스와 음성인식 키보드를 제어할 수 있다. 기존의 이러한 입력장치를 사용할 수 없는 산업 분야에서 활용할 수 있도록 하였다. 비접촉 방식의 입력장치 프로그램이기 때문에 메타버스와 의료 산업 분야에서 활용성이 높을 것으로 생각된다.

II. Preliminaries

본 논문의 알고리즘은 OpenCv-Python을 이용하여 실시간 영상 처리를 토대로 만들어졌으며 MediaPipe에서 제공하는 손가락 마디의 랜드마크를 사용한다. Pyautogui와 Pynput 라이브러리를 통해 키보드와 마우스를 제어하며, Pyautogui의 경우 파이썬에서 한글 키를 사용할 수 있도록 구현하기 위해서 사용했다. Pillow 라이브러리를 이용하여 한글을 처리할 수 있게 구현하였으며, Speech_recognition

을 사용하여 음성인식 처리를 구현하였다. 사용 언어로는 Python-3.8 버전을 사용하였으며, 사용 가능한 OS로는 Windows10과 Windows11이 있다.

III. The Proposed Scheme

키보드 기능의 경우 MediaPipe에서 제공하는 손가락 마디 랜드마크 중 4, 8번 즉, 검지의 끝부분과 엄지의 끝부분을 사용한다. 랜드마크를 통해 x, y 좌표를 받고 엄지와 검지 부분의 x, y 좌표를 뺀 절대값을 통해 엄지와 검지 사이의 거리를 알아낼 수 있다. 이를 통해 엄지와 검지 사이가 일정 이상 붙게 되면 키보드 입력 이벤트를 발생시키도록 구현하였으며 실시간 영상 처리로 키보드 사용자 인터페이스와 기능 변환 인터페이스를 구현하였다.

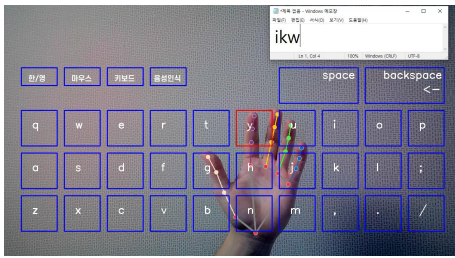


Fig. 1. Enter Keyboard

마우스 기능은 키보드와 달리 중지의 랜드마크를 사용한다. 중지의 랜드마크 x, y 좌표를 받아 각 좌표를 화면의 크기와 1:1 대응으로 마우스를 제어할 수 있도록 구현하였으며 마우스 클릭 구현은 키보드 제어와 같은 알고리즘을 적용하여 중지와 검지가 일정 이상 붙게 되면 클릭하는 이벤트를 발생시키도록 구현하였다.

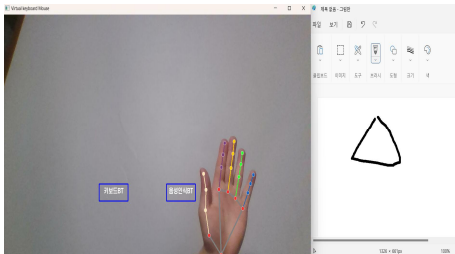


Fig. 2. Mouse control

음성인식 키보드의 음성인식 구현 부분은 구글에서 제공하는 오픈 소스와 데이터를 사용하였으며, 본 논문의 프로그램과 결합하여 음성 인식 프로그램 켜짐 상태가 되면 2~3초간 음성 입력 신호를 버퍼에 저장한 뒤 어떤 소리도 들리지 않으면 버퍼에 저장된 음성 입력 신호를 커서에서 글자 형태로 출력되도록 하였다.

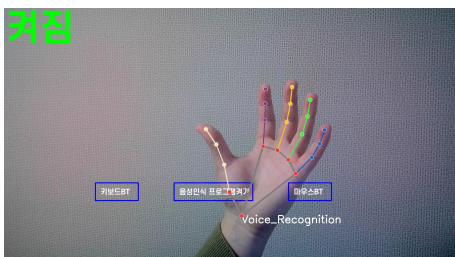


Fig. 3. Speech recognition

Table 1. Experiment environment

CPU	i5-1135G7
RAM	16GB
OS	Windows 11

실시간 영상 처리를 통해 두 손을 인식하여 두 손으로 키보드 기능을 사용할 수 있게 하였으며 키보드 기능에서 검지가 가리키는 부분은 빨간색으로 재처리하여 어떤 것을 가리키고 있는지 명확하게 하였다. 키보드 입력 이벤트는 벡터를 이용한 엄지, 손목, 검지의

각도를 계산해 그 값을 통해 제스처를 학습시켜 입력 제스처로 활용하는 방법을 제안하였지만, 학습 데이터가 많아짐에 따라 끊김 현상이 발생하였으며, 본 연구에 적용된 방법보다 연속적인 동작 실행에 있어서 인식이 저조하였다. 따라서 x, y 좌표를 통한 이벤트 발생 방법으로 키보드 입력 및 기능 전환 제어 방법을 결정하였다. 향후 입력 이벤트는 개선과 보완의 여지가 있을 것으로 사료된다.

IV. Conclusions

본 연구의 추후 활용방안으로는 코로나 시대에 따라 VR/AR 기기가 일반 가정에 보급이 되고 있는데 이때 사용되는 센서는 고가의 센서를 사용하기 때문에 웹캠과 프로그램으로 작동되는 본 논문의 프로그램이 비용 절감 측면에서 장점을 보일 것으로 기대된다. 또한, 키 매핑이 가능하기 때문에 키오스크, 자동차 등 여러 산업분야에서 응용하여 사용할 수 있을 것으로 추측한다. 의료 산업 분야에서 기존의 접촉 방식의 키보드, 마우스 보다 비접촉식의 본 연구에서 제안한 방법을 사용한다면 접촉으로 인한 감염 질환을 예방할 수 있을 것으로 사료된다.

REFERENCES

- [1] <https://github.com/gadeokexp/VoiceTyper>
- [2] https://github.com/yasser-sulaiman/virtual_keyboard