AI를 활용한 손가락 인식 및 가상 터치 서비스

조아라[1], 유승배[1], 윤병훈[1], 조형주[2], 하광림[3]
[1]경북대학교 소프트웨어학과 학부생
[2]경북대학교 소프트웨어학과 교수
[3](주)씨에스리 AI엔지니어링사업부 사업부장
zoanna5442@gmail.com, {e10013, gland45}@naver.com, hyungju@knu.ac.kr, dilector@naver.com

Finger Recognition and Virtual Touch Service using AI

A-Ra Cho[1], Seung-Bae Yoo[1], Byeong-Hun Yun[1], Hyung-Ju Cho[1], Gwang-Rim Ha[2]
[1]Dept. of Software, Kyungpook National University
[2]AI Engineering Division, CSLEE. CO., LTD.

요 약

코로나-19로 인해 비접촉 서비스의 중요성이 더욱 대두되고 있다. 키보드나 마우스와 같은 기존 입력 장치를 대체하기 위해 사람들은 디지털 기기에서 손을 사용하여 자연스럽고 간단한 입력을 할 수 있게 되었다. 본 논문에서는 미디어파이프(MediaPipe)와 LSTM(Long Short-Term Memory) 딥러닝을 활용하여 손 제스처를 학습하고 비접촉 입력 장치로 구현하는 방법을 제시한다. 이러한 기술은 가상현실(VR; Virtual Reality), 증강현실(AR; Augmented Reality), 메타버스, 키오스크 등에서 활용 가능성이 크다.

1. 서론

기존의 가상 터치 기술은 3D 카메라나 적외선 카메라와 같은 정밀한 센서가 장착된 장비가 필요하다. 이로 인해 제조 비용이 증가하여 가상 터치 기술을 보급하는 데 어려움이 있다. 본 논문은 위와 같은 문제를 해결하고자 일반 카메라를 사용하여 비접촉 입력 장치를 구현하였다. 젯슨 나노(Jetson Nano)의 CSI 광각 카메라와 LSTM (Long short-term memory) 딥러닝 알고리즘을 사용하여 가상 터치를 구현한다. 결과적으로 사용자 친화적인 제스처를 동적 제스처 학습을 통해 인식률을 높이고, 추가 보정을 통해 클릭 위치를 수정했다.

2. 관련연구

본 연구는 구글(Google)이 제스처 인식을 위해서 공개한 프레임워크인 미디어파이프(MediaPipe)를 사용했다[1]. 미디어파이프에서 제공하는 손 모델을 사용하면 손 관절에 해당하는 21개의 특징점과 각 특징점의 좌표를 알 수 있다. 딥러닝에서 시계열 데이터를 학습시키는 대표적인 모델은 순환신경망(Recurrent Neural Network)이다. 순환신경망의 기울기 소멸 문제를 해결하기 위해 LSTM 모델이 제안되었다. 본 논문에서는 LSTM 모델을 이용하여 동적 제스처를 인식하도록 설계하였다.

3. 연구 방법

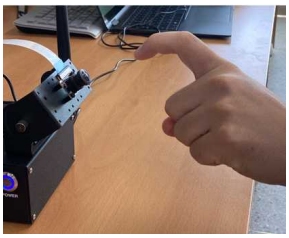
일반 카메라를 사용한 기존 연구에는 다음과 같은 문제점들이 있다. 첫째, 미디어파이프를 활용하여 손끝의 거리를 계산하여 인터페이스를 제어한다[2]. 이 같은 경우 사용자가 익숙하지 않은 특정 제스처를 학습해야 한다. 둘째, 제스처 모델을 CNN(Convolutional Neural Network)으로 학습하여 단일 손가락 모델을 생성하였다[3]. CNN 방식으로 정적 제스처를 학습한 모델은 사용자가 특정 각도에서 손 모양을 만들 때만 제스처가 인식되는 문제가 있다.

본 연구는 사용자 친화적으로 '검지를 구부리는 제스처'를 클릭으로 인식하도록 학습하여 구현하였다. 또한 손가락을 다양한 각도로 구부렸을 때 제스처를 인식할 수 있도록 LSTM 딥러닝 기법을 사용하였다. LSTM 학습 시에는 미디어파이프를 사용해 추출한 특징점들의 각도를 입력데이터로 사용하였다.

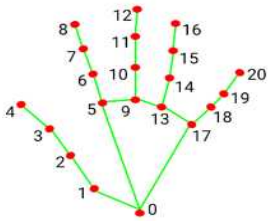
4. 구현

본 연구에서는 젯슨 나노를 클라이언트로 사용했다. 손은 젯슨 나노에 장착된 CSI 광각 카메라를 사용하여 촬영하였다. 또한 젯슨 나노에 설치된 미디어파이프 프레임워크의 손 모델을 활용하여 손 객

체 인식을 수행하였다. 손 동작 정보를 정확하게 인식하기 위해 관절 지점 사이의 각도 정보를 활용하여 제스처를 인식한다.



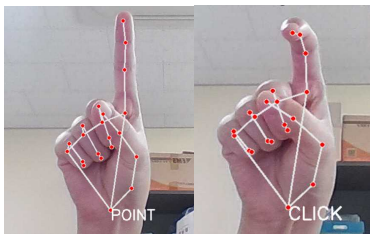
(그림1. 젯슨나노)



(그림2. 랜드마크)

제스처 동작은 포인트(Point) 동작과 클릭(Click) 동작으로 지정하였다. 한 동작에 대해 2초짜리 비디오 클립을 직접 생성했다. 그림3은 직접 데이터를 생성하는 사진이다. 각각 500개의 비디오 클립을 생성하여 총 1,000개의 데이터셋으로 학습하였다. 관절점 기반으로 관절 사이의 각도를 계산하여 입력 벡터를 구성한다. 한 동작을 60개의 프레임으로 촬영하고 각 프레임에서 계산한 입력 벡터를 순차적으로 장단기기억 모델에 입력하여 학습시킨다.

실험 데이터셋은 운영체제는 윈도우11, CPU는 Intel i9, GPU는 NVIDIA GeForce RTX 2080 SUPER 환경에서 Python 3.11.2와 OpenCV 4.7.0 프로그램을 사용하여 생성하였다.



(그림3. 손동작 학습 데이터셋)

본 논문에서 사용된 학습 모델은 하나의 LSTM 레이어와 두 개의 완전 연결 레이어로 구성했다. 학습 결과, 손실률은 0.0211, 정확도는 0.9931로 정확도 측면에서 높은 것으로 나타났다. 학습 모델은 AWS(Amazon Web Service) 서버에 올려 제스처를 예측하는 데 사용한다.

손 관절 사잇각 정보를 클라이언트에서 서버로 실시간 전송한다. 수신받은 데이터를 사용하여 모델이 동작에 대한 예측값을 반환한다. 예측값이 더 높은 동작 데이터를 클라이언트로 전송한다. 클라이언트는 전송받은 결과에 따라 9번 좌표에서 파이썬의 파이앤풋(pynput)모듈을 사용해 이벤트를 처리한다.

성능 향상을 위해 텐서플로(Tensorflow) 모델 최적화 툴킷(toolkit)과 텐서플로 경량화 기법을 사용하여 추론 최적화와 모델 경량화를 진행하였다.

(그림4. 핸드트래킹과 클릭시 프로그램 실행)

5. 결론

본 논문은 LSTM 모델을 활용하여 3차원 가상 공간에서 손 제스처를 인식하고 화면상에서 마우스를 조작하는 방법을 제시한다. 제스처 데이터는 일반 사용자 동작을 기반으로 구성되었다. 본 연구를 기반으로 하여 일반 카메라와 소프트웨어를 활용하여 고비용 카메라를 사용하지 않고 응용 분야를 확대할 수 있다. 현재 이 기술로 다수의 카메라로 한 개의 키오스크를 제어하는 데 활용할 수 있어 비용 절감의 이점을 얻을 수 있다. 제안된 기법은 가상현실, 증강현실, 메타버스 등에서 다양하게 활용될 수 있다.

감사의 글

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1]김신영, 엄서정, 유선영, 김수정, 이경미, “미디어파이프와 장단기기억을 이용한 수화 동작인식 앱 개발”, 한국디지털콘텐츠학회논문지, 24(1), pp. 111-119, 2023.

[2]김유라, 서솜이, 박구만, "제스처 인식을 이용한 비접촉식 미디어 아트 키오스크", 한국방송미디어공학회 학술발표대회, 서울, 2022, pp. 232-235.

[3]허경용, 송복득, 김지홍, “손 표현 인식을 위한 계층적 손 자세 모델”, 한국정보통신학회논문지, 25(10), pp. 1323-1329, 2021.