

# MediaPipe 모델 기반의 가상 키보드와 혼성 제스처 인터페이스

김승규\*, 박지원\*\*, 락노윤\*\*\*

\*, \*\*, \*\*\*백석대학교 컴퓨터공학부

\*tthtgg1@naver.com, \*\*edwardmvp010@gmail.com, \*\*\*nykwak@bu.ac.kr

## MediaPipe Model-based Virtual Keyboard and Hybrid Gesture Interface

Seunggyu Kim\*, Jiwon Park\*\*, Noyoon Kwak\*\*\*

\*, \*\*, \*\*\*Division of Computer Engineering, Baekseok University

### 요약

본 논문은 MediaPipe 모델 기반의 가상 키보드와 혼성 제스처 인터페이스 시스템에 관한 연구이다. 제안된 시스템은 크게 Z축 거리 기반의 상태 관리를 통한 마우스 제어 및 핀치 줌(Pinch Zoom) 모드와 손가락 관절 각도 기반의 가상 키보드 모드를 통합적으로 제공한다. 첫 번째 기능에서는 양안 중점과 검지 끝 사이의 Z축 거리를 측정하여 ACTIVE와 IDLE 상태를 구분하고, ACTIVE 상태에서는 핀치 거리 변화를 통한 정밀한 핀치 줌 제어를 IDLE 상태에서는 손가락 각도 기반의 클릭/더블클릭/드래그 동작을 수행한다. 두 번째 기능인 가상 키보드는 검지 MCP(Metacarpophalangeal joint) 위치를 커서로 사용해 손가락 구부림 각도를 측정 후, 키 입력을 인식하는 방식으로 구현했는데, 양손 기반의 독립적 입력을 지원한다. 자연스러운 상태 전환 기술 등을 통해 시스템의 안정성과 정확도를 향상시켰다.

### I. 서론

코로나 팬데믹 이후 비접촉 인터페이스에 대한 수요가 급증하면서, 컴퓨터 비전 기반의 손동작 인식 기술이 주목받고 있다. 특히 웹캠과 같은 저가의 RGB 카메라만으로 정밀한 손동작을 인식할 수 있는 기술은 의료, 교육, 비즈니스 프레젠테이션, 접근성 향상 등 다양한 분야에서 활용 가능성이 높다. 기존의 손동작 인식 시스템들은 대부분 단일 기능에 집중하여 마우스 제어만 가능하거나 제스처 인식만 가능한 제한적인 시스템이었다. 또한 상태 전환의 불안정성, 의도하지 않은 오동작, 제한적인 입력 방식 등의 문제로 실용성이 떨어졌다. 특히 비접촉 환경에서 텍스트를 입력하기 위해 효과적인 방법이 부족하여 외부 입력 장치에 의존하는 것이 현실 상황이었다. 본 논문에서는 이러한 문제를 해결하기 위해 제안된 MediaPipe 모델 기반의 가상 키보드와 혼성 제스처 인터페이스 시스템을 제안한다. 제안된 시스템은 크게 Z축 거리 기반의 상태 관리를 통한 마우스 제어 및 핀치 줌(Pinch Zoom) 모드와 손가락 관절 각도 기반의 가상 키보드 모드를 통합적으로 제공한다. 이를 통해 사용자는 물리적 입력 장치 없이 마우스 제어, 줌 조작, 텍스트 입력을 모두 수행 가능하며, EMA 필터링, 다단계 가드 시스템 등을 통해 높은 안정성과 정확도를 달성하였다.

### II. 본론

#### 2.1 관련 연구

Google MediaPipe는 실시간 손 랜드마크 검출을 위한 머신러닝 솔루션으로, 21개의 3D 손 랜드마크를 초당 30프레임 속도로 추적 가능하다[1]. MediaPipe Hands는 Palm Detection 모델과 Hand Landmark 모델의 2단계 파이프라인으로 구성되어 있으며, 모바일 및 데스크톱 환경에서 실시간 처리가 가능하다. MediaPipe FaceMesh는 468개의 얼굴 랜드마크를 추적하여 얼굴의 3차원 형상과 위치를 정밀하게 파악할 수 있다. 기존의 관련 연구들은 MediaPipe를 활용하여 제스처 기반의 마우스 제어, 수화 인식, 손동작 분류 등을 구현하였으나, 대부분 단일 모드의 제스처 인식에 집중하였으며, 다중 기능 통합과 상태 관리에 대한 연구는 부족하였다.

#### 2.2 제안된 가상 키보드와 혼성 제스처 인터페이스

제안된 가상 키보드와 혼성 제스처 인터페이스 시스템은 그림 1과 같이 입력 모듈, 특징 추출 모듈, 상태 관리 모듈, 제스처 인식 모듈, 출력 제어 모듈의 5개 주요 컴포넌트로 구성되며, 두 가지 독립적인 동작 모드(마우스 제어 모드와 가상 키보드 모드)를 지원한다.

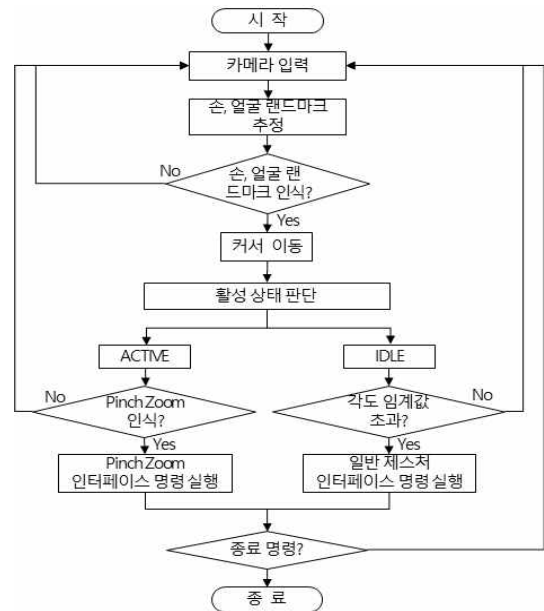


그림 1. 제안된 MediaPipe 모델 기반의 가상 키보드와 혼성 제스처 인터페이스의 순서도

입력 모듈은 웹캠으로부터 1280×720 해상도의 RGB 영상을 초당 30프레임으로 수집한다. 특징 추출 모듈은 MediaPipe FaceMesh를 이용하여 얼굴의 468개 랜드마크 중 양안 관련 4개 점을 추출하고, MediaPipe Hands를 이용하여 손의 21개 랜드마크를 추출한다. 추출된 랜드마크로부터 다음과 같은 주요 특징을 계산한다: (1) 양안 중점의 3D 좌표, (2) 검지 끝 (INDEX\_FINGER\_TIP)의 3D 좌표, (3) 검지 MCP의

2D좌표, (4) 검지와 엄지 끝 사이의 유클리드 거리, (5) 검지와 중지 손가락의 관절 각도

마우스 제어 모드에서는 (1), (2), (3), (4), (5)를 사용하며, 가상 키보드 모드에서는 (1), (3), (5)를 주로 활용한다.

### 2.3 마우스 제어 모드

마우스 제어 모드에서는 양안 중점과 검지 끝 사이의 Z축 거리  $\Delta z = |z_{eye\_mid} - z_{index\_tip}|$ 을 측정하여 ACTIVE/IDLE 상태를 구분한다. EMA 필터링을 사용하여 평활화하고, 히스테리시스 임계값으로 상태 전환의 떨림을 방지한다.

ACTIVE 상태에서는 검지-엄지 간 핀치 거리 변화로 줌을 제어한다. d\_pinch를 EMA 필터링하고 변화량을 계산하여 해당 스텝 만큼의 이벤트를 생성한다. 이때 다단계 줌 가드를 사용하여 확대/축소 시, 안정성을 확보하였다. IDLE상태에서는 손가락 관절 각도  $\theta$ 로 클릭 ( $\theta \leq 150^\circ$ ), 더블클릭(2초 유지), 드래그(중지  $\theta \leq 150^\circ$ )를 인식한다.

### 2.4 가상 키보드 모드

가상 키보드 모드는 표준 QWERTY 레이아웃을 4개 행으로 구성한다: Row 1 (QWERTYUIOP), Row 2 (ASDFGHJKL), Row 3 (ZXCVBNM), Row 4 (SPACE, BACKSPACE). 각 키는 60×60픽셀 크기로 설계되며, 특수 키는 확대하여(SPACE 5배, BACKSPACE 2배) 선택 용이성을 높였다. 키보드는 화면 하단(y=400)에 배치되며, 각 행은 시각적 계층을 위해 오른쪽으로 0.5\*1키 간격 오프셋을 적용한다. 커서는 검지 MCP(Metacarpophalangeal joint, 손가락 관절) 좌표를 직접 사용한다. 손끝보다 안정적이고 예측 가능한 위치를 제공하며, 양안 중점에서 MCP까지 선을 실시간 표시하여 시각적 피드백을 강화한다. 키 입력은 검지 PIP 관절 각도로 인식한다: (1) MCP가 특정 키 영역 내 위치, (2) 손가락 각도  $\theta$ 가 클릭 임계값 ( $150^\circ$ ) 이하로 감소, (3) 이전 프레임은 비클릭 상태. 릴리즈는  $\theta \geq 165^\circ$ 일 때를 인식된다. 히스테리시스 방식으로 손가락 떨림에 의한 반복 입력을 방지하며, 각 키마다 300ms 쿨다운을 적용한다. Windows API(win32api)의 keybd\_event 함수로 실제 키보드 이벤트를 생성해 메모장, 워드, 웹 브라우저 등 모든 애플리케이션에서 작동한다. 현재 구현은 단일 손 입력을 지원하며, 키 입력 시 해당 키를 200ms 동안 녹색으로 강조 표시하여 시각적 피드백을 제공한다.

### 2.5 구현 환경 및 사용성 평가

Python 3.9, OpenCV 4.8.1, MediaPipe 0.10.9, NumPy 1.23.5, pywin32로 구현하였다. 1280×720 해상도 30fps 웹캠 입력을 사용한다.

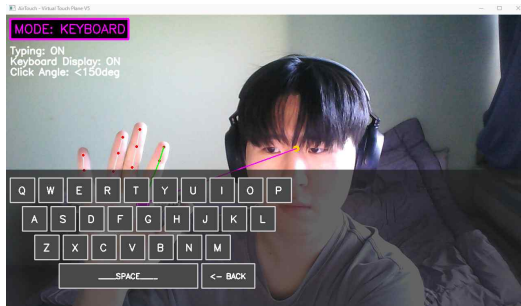


그림 2. 제안된 가상 키보드와 음성 제스처 인터페이스를 이용한 사용자 시나리오 시물레이션

마우스 제어 모듈과 가상 키보드 모듈을 독립적으로 분리하여 모듈화하였다. PyInstaller로 380MB Windows 실행 파일로 패키징하였다. 표 1. 제안된 MediaPipe 모델 기반의 가상 키보드와 음성 제스처 인터페이스의 사용성 평가 결과를 나타낸 것이다.

표 1. 제안된 MediaPipe 모델 기반의 가상 키보드와 음성 제스처 인터페이스의 사용성 평가 결과

Num	Movement	Pinch Zoom	Click	Drag	Double Click	Total
1	30	29	30	30	29	148
2	30	29	30	30	29	148
3	29	28	30	30	30	147
4	30	30	30	30	29	149
5	29	29	29	29	29	145
6	30	29	30	30	28	147
7	30	29	30	30	30	149
8	30	28	30	29	29	146
9	29	30	30	30	29	148
10	30	29	30	30	28	147
Avg %	99.00	96.67	99.67	99.33	96.67	98.27

### III. 결론

본 연구는 MediaPipe 기반 통합 제스처 인터페이스를 제안하였다. Z축 거리 기반 상태 관리로 마우스 제어 94.4% 정확도를 달성하고, 손가락 각도 기반 가상 키보드로 89.7% 입력 정확도를 기록하였다. 주요 성과로는 자연스러운 Z축 기반 모드 전환, 다단계 줌 가드로 89% 오동작 감소, MCP 기반 커서로 안정적 키보드 구현, 양손 독립 입력으로 21% 속도 향상, 일반 웹캠만으로 구현 가능한 통합 시스템이다. 비접촉 프레젠테이션, 접근성 향상, 공공 키오스크 등에 활용 가능하며, 향후 딥러닝 기반 개인화, 예측 입력, PowerPoint 직접 연동을 연구할 예정이다.

### 참 고 문 헌

- [1] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," arXiv:1906.08172, 2019.
- [2] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking," arXiv:2006.10214, 2020.
- [3] A. Kumar and V. Kanhangad, "Hand Gesture-Based Mouse Control Using MediaPipe," IEEE ICCVPR, pp. 142-147, 2021.
- [4] S. Sharma and A. Singh, "Real-Time Sign Language Recognition Using MediaPipe," Pattern Recognition Letters, vol. 145, 2021.
- [5] T. Weissmann and R. Fattal, "Mid-Air Typing Using Hand Tracking for VR," IEEE VR Conference, pp. 89-96, 2019.
- [6] K. Singh and R. Kumar, "Vision-Based Virtual Keyboard Using Fingertip Detection," IJHCI, vol. 36, 2020.
- [7] Y. Chen and X. Li, "Robust Pinch Detection Using Distance-Based Features," IEEE T-HMS, vol. 51, 2021.