



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

모음 손동작 분리를 이용한 가상
키보드 입력 속도 개선

Fast Virtual Keyboard Typing Using Vowel
Hand Gesture Recognition

2023 년 2 월

서울대학교 대학원

전기 정보 공학부

박수진

모음 손동작 분리를 이용한 가상 키보드 입력 속도 개선

지도 교수 이혁재

이 논문을 공학석사 학위논문으로 제출함
2023 년 2 월

서울대학교 대학원
전기 정보 공학부
박수진

박수진의 공학석사 학위논문을 인준함
2023 년 2 월

위원장 _____ 조남익 _____ (인)

부위원장 _____ 이혁재 _____ (인)

위원 _____ 이태호 _____ (인)

초 록

최근 AR(Augmented Reality) 및 VR(Virtual Reality)기술이 발전함에 따라 사용자 인터페이스 기술에 대한 연구가 활발히 진행되고 있다. 키보드, 마우스, 터치스크린 등 기존의 입력 장치를 대체할 수 있는 가상 입력 인터페이스로 손동작 인식이 주목받고 있으며, 이에 다양한 접근 방식으로 손동작 인식 모델이 제안되었다. 정교한 손 인식을 위해 센서가 내장된 장치형 인터페이스를 사용한 손동작 인식 연구가 이뤄지고 있다. 최근에는 딥러닝 기술의 발전에 따라 별도의 장치가 필요 없는 카메라 기반 비전 인식을 이용한 손동작 인식이 AR/VR의 핵심 어플리케이션인 가상 키보드, 가상 글쓰기, 수화 인식 등에 사용되기 위한 인터페이스로 활발한 연구가 이뤄지고 있으며, 특히 객체 인식 딥러닝 네트워크의 경우 실시간으로 손동작을 인식할 수 있다는 점에서 많이 사용되고 있다.

본 논문에서는 알파벳의 사용 빈도 및 연속적으로 나타나는 두 문자 간의 상관 빈도에 대해 통계적 분석을 진행 시, 모음의 사용 빈도수는 평균 39%이며 자음-모음 또는 모음-자음 순으로 배치되는 경우는 약 65% 정도로 모음의 입력의 속도를 향상시켜 가상 키보드 입력 속도를 개선하는 방법을 제안한다.

제안한 방식은 객체 인식 딥러닝 네트워크를 이용하여 가상 키보드 입력에 필요한 6개의 손동작에 5개의 알파벳 모음, ‘a’, ‘e’, ‘i’, ‘o’, ‘u’에 대해 특정된 손동작을 추가로 학습시켜, 자음의 경우 기존과 동일하게 가상 키보드 배열 내에서 선택되어 입력되도록 하며 모음의 경우 모음 손동작이 인식되면 입력되도록 한다. 인식이 필요한 손동작이 늘어남에 따라 자동 라벨링을 통해 다수의 학습 데이터를 확보하여 94.39%의 비교적 높은 평균 손동작 인식률을 확보했다. 제안한 방식 기반의 가상 키보드를 구현하여 사용자 테스트를 진행한 결과 기존에 제안된 가상 키보드 입력 방식보다 제안한 방식의 속도가 12.45% 개선되었다.

주요어 : 가상 키보드, 가상현실, 증강현실, 손동작 인식, 객체 인식, HCI(Human Computer Interaction)

학 번 : 2021-20860

목 차

제 1 장 서 론.....	1
제 1 절 연구 배경 및 내용.....	1
제 2 절 논문의 구성.....	3
제 2 장 관련 연구.....	4
제 1 절 손동작 인식 관련 연구.....	4
제 2 절 가상 키보드 관련 연구.....	7
제 3 장 가상 키보드 속도 향상 방안 제안.....	9
제 1 절 모음 손동작 추가 배경.....	9
제 2 절 제안한 가상 키보드 입력 시스템.....	10
제 4 장 실험 결과 및 분석.....	14
제 1 절 입력 시간 예측 시뮬레이션 결과.....	14
제 2 절 딥러닝 네트워크의 학습.....	15
제 3 절 가상 키보드의 사용자 실험 및 분석.....	18
제 5 장 결론.....	20
참고문헌.....	21
Abstract.....	25

표 목차

[표 3-1] 모음과 자음의 사용 빈도수.....	9
[표 4-1] 가상 키보드 입력 속도 simulation 결과.....	15
[표 4-2] 사용자 실험 문장.....	19

그림 목차

[그림 1-1] 손동작으로 증강 현실을 조작하는 모습(영화 마이너 리티 리포트 포스터).....	2
[그림 2-1] 장갑 형태의 센서.....	4

[그림 2-2] MYO Arm band	5
[그림 2-3] YOLO 모델	6
[그림 2-4] YOLO 네트워크의 구조.....	6
[그림 2-5] 추론 시간과 mAP (IOU@0.5).....	6
[그림 2-6] 클릭 손동작의 과정.....	8
[그림 2-7] Multy-layer 손동작.....	8
[그림 3-1] 알파벳 상관 빈도수.....	10
[그림 3-2] 모음 손동작	10
[그림 3-3] 제안한 입력 시스템의 손인식 흐름도.....	11
[그림 3-4] 가상 키보드 모음 입력 예시	12
[그림 3-5] 가상 키보드 자음 입력 예시	13
[그림 4-1] layer-aware 키보드 레이아웃	15
[그림 4-2] 제안한 키보드 레이아웃.....	16
[그림 4-3] 자동 라벨링 흐름도.....	17
[그림 4-4] 자동 라벨링으로 생성된 학습 데이터셋 예시	18
[그림 4-5] 클래스 별 손동작 인식률	18
[그림 4-6] 사용자 실험 결과.....	20

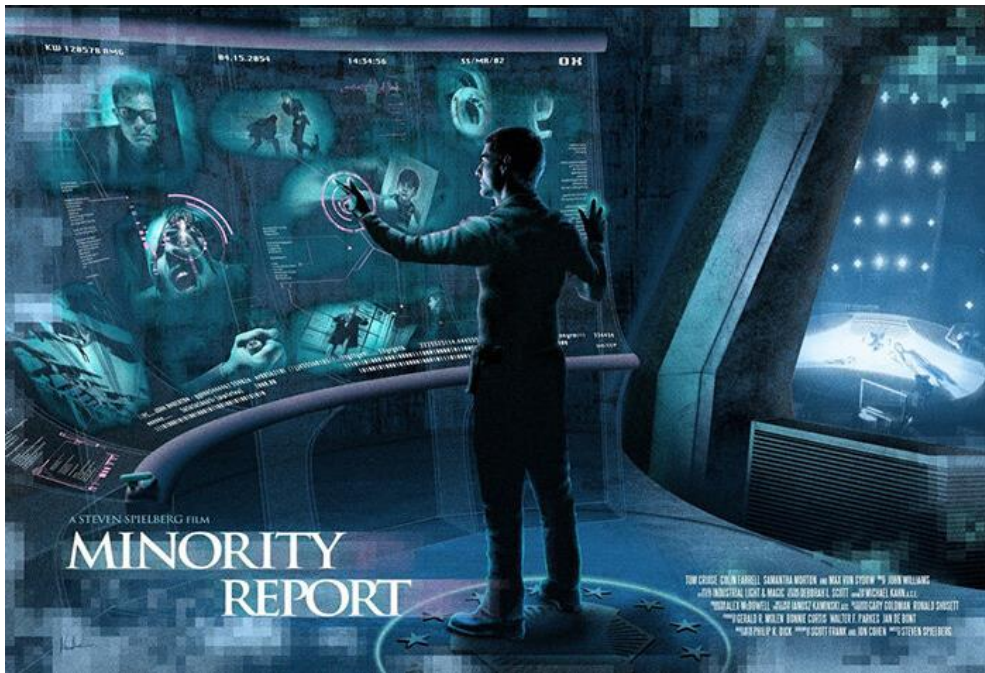
제 1 장 서 론

제 1 절 연구 배경 및 내용

최근 AR(Augmented Reality) 및 VR(Virtual Reality)기술이 발전함에 따라 사용자 인터페이스 기술에 대한 연구가 활발히 진행되고 있다 [1-3]. 그 중 손동작 인식은 가상 키보드, 가상 글쓰기, 수화 인식 등 많은 AR/VR 애플리케이션의 핵심기술이다. 손동작 인식을 위해 센서를 손과 손등에 부착하여 센서의 신호를 탐지해 손의 움직임을 탐지하는 인터페이스[4], 손에 색상이 있는 점들을 표시하여 점들의 움직임을 통해 손동작을 인식하는 마커 기반 인터페이스[5][6], 센서가 장착된 장갑을 착용하여 손의 움직임을 간접적으로 감지하는 방식의 인터페이스[7] 등 장치 기반의 손동작 인식 인터페이스에 대한 연구가 활발히 이뤄지고 있다. 하지만 이와 같은 방식들은 AR/VR 기기 외 별도의 장치가 필요없이 사용자가 자유롭게 움직일 수 있고 가상세계에서도 편리하게 이용 가능한 인터페이스는 아니다. 이러한 문제를 해결하기 위해 장비를 사용하지 않고 손을 인식할 수 있는 연구가 다양한 접근법으로 진행되고 있다. Xu et al.[8]은 RGBD 카메라 이미지 입력에 따른 3D 손동작 인식 모델을 제안하였다. Lee et al. [9]은 손의 피부색 정보를 이용해 손 영역을 분리한 이후 외곽선을 처리하여 손동작 및 손가락 위치를 인식하는 방법의 인터페이스를 제안하였다. Ekhbat은 CNN(Convolutional Neural Networks) 기반의 다양한 환경에서도 손동작을 인식할 수 있는 인터페이스를 제안하였다[10]. 실제 AR/VR 환경에 적용되기 위해서는 실시간으로 손동작 인식이 되어야 하며 높은 인식률을 가져야 하기 때문에 최근에는 객체 인식 기반 딥러닝 네트워크를 사용한 손동작 인식 인터페이스 연구가 활발히 진행 중이다. 이러한 인터페이스는 특히나 기존의 키보드, 마우스, 터치스크린과 같은 입력 인터페이스를 대체할 수 있을 것으로 기대되고 있으며 가상 키보드가 그 대표적인 예이다. 가상 키보드가 실용적으로 사용되기 위해서는 정교하고 빠른 손동작 인식뿐 아니라 입력 속도의 개선을 통해 실제 사용 환경에서의 효용성 확보가 중요하다.

본 논문에서는 객체 인식 기반 딥러닝 네트워크 기반의 손동작 인식을 활용한 가상 키보드 입력 시스템을 통해 모음 손동작을 추가하여 가상 키보드 입력 속도를 높이는 방식을 제안한다. 알파벳의 사용

빈도와 연속적으로 나타나는 두 문자 간의 상관 빈도 통계적 분석을 진행한 결과, 26개의 알파벳 가운데 약 19%의 비중을 차지하는 모음의 사용 빈도는 평균 39%이며, 모음 다음에 자음 또는 자음 다음에 모음이 번갈아 나오는 상관 빈도는 약 65%의 비중을 차지했다. 이를 근거로 하여 모음 입력 시 각 모음에 대해 특정된 손동작을 인식하여 입력할 수 있도록 하여 입력 속도를 개선한 가상 키보드를 제안한다. 성능 분석을 위해 가상 키보드의 입력 시간 예측을 위한 시뮬레이션 모델을 고안하여 기존에 제안된 가상 키보드 시스템에서의 입력 속도를 비교한다. 마지막으로 실제 가상 키보드 상에서 사용자 실험을 통해 제안한 방식의 효용성을 증명한다.



[그림 1-1] 손동작으로 증강 현실을 조작하는 모습(영화 마이너리티 리포트 포스터)

제 2 절 논문의 구성

본 논문의 구성은 다음과 같다. 2장에서는 손동작 인식 기술, 손동작 인식 기반의 가상 키보드 연구 및 객체 인식 딥러닝 네트워크를 활용한 가상 키보드에 대해 이전에 연구된 내용을 소개한다. 3장에서는 모음 손동작 추가를 통해 모음과 자음의 입력 방식을 구별하여 객체 인식 기반 딥러닝 네트워크를 사용한 가상 키보드 입력 속도를 개선할 수 있는 방식을 제안한다. 4장의 1절에서는 시뮬레이션 모델을 이용한 입력 속도 개선을 예측하고 2절에서는 제안한 방식으로 손동작 인식을 진행하였을 때의 인식률을 확인한다. 3절에서는 제안한 알고리즘 기반 가상 키보드를 구현하여 사용자 실험을 진행하여 결과를 분석한다. 마지막 장에서는 본 논문의 결론을 내리도록 한다.

제 2 장 관련 연구

제 1 절 손동작 인식 관련 연구

손동작 인식은 특정한 장치나 센서를 이용하여 인식하는 방법과 카메라를 입력 장치로 하여 비전 인식 기반의 방법으로 나누어 생각해볼 수 있다.

전자의 경우 대표적인 예로는 [그림 2-1]의 장갑 형태의 인터페이스로 장갑에 부착된 다수의 센서를 이용하여 손가락의 좌표, 그리고 손의 움직임을 감지할 수 있다. 이 외에도 [그림 2-2]의 MYO armband의 경우 EMG(Electromyography) 센서와 IMU 센서를 내장하여 근육이 움직이면서 발생하는 전기신호를 통해 손동작을 분석한다[11]. 그러나 이런 방법들은 사용자가 장갑 또는 암밴드와 같은 별도의 장치를 가지고 이동해야 한다는 점에서 사용자에게 불편함을 야기할 수 있으며, 이러한 장치를 구매하기 위한 추가적인 비용이 발생한다는 단점이 있다.

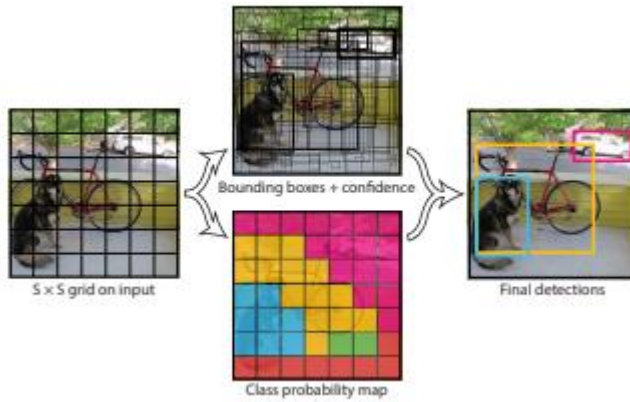


[그림 2-1] 장갑 형태의 센서

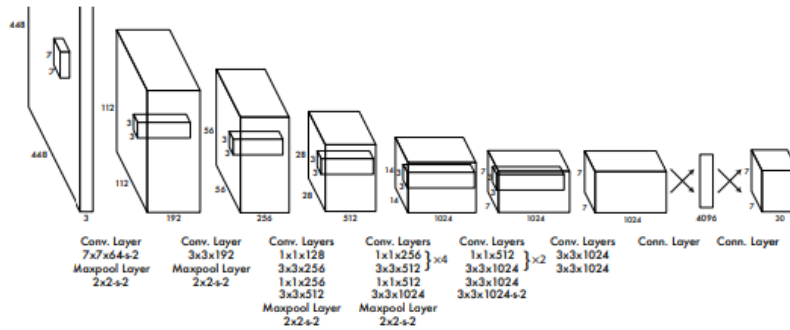


[그림 2-2] MYO armband

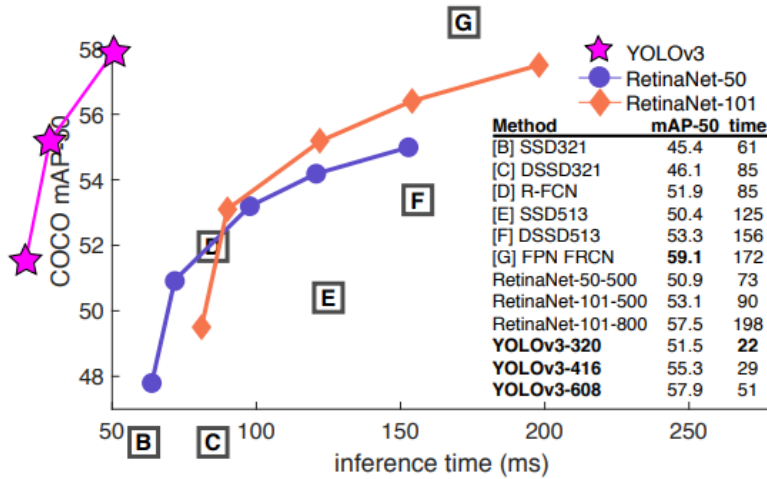
후자의 경우 사용자에게 움직임 상의 편의를 제공할 수 있다는 장점을 가진다. 카메라 기반 비전 인식 관련 Elmezain et al.[12]은 HMM(Hidden Markov Model)을 이용한 시계열 데이터 분석 및 피부색과 3D depth map을 통해 손을 감지하여 손의 궤적 정보를 확보하여 손의 움직임을 추적하는 방식을 제안하였다. Xu et al.[13]은 손색상 정보를 이용하여 RGB 이미지 중 손을 포함한 ROI를 추출하여 딥러닝 네트워크를 이용해 손동작을 분류하는 방식을 제안하였다. 나아가 최근 딥러닝의 발전에 따라 실시간으로 사물을 인식하고 높은 인식률을 가지는 R-CNN(Region-CNN) [14], SSD(Single Shot MultiBox Detector) [15] 과 YOLO(You Only Loo Once) [16]과 같은 객체 인식 딥러닝 네트워크를 활용하여 손 영역 추출 및 손동작을 인식하는 방식도 제안되었다. 이 가운데 YOLO는 객체의 분류와 영역 추출을 동시에 해결하는 1-stage detector 방식을 처음으로 고안해서 실시간 객체 검출을 가능하게 한 모델이다. [그림 2-3]과 같이 SxS 그리드 영역으로 나누어진 이미지들이 [그림 2-4]의 컨볼루션 신경망을 통과하고 나면 각 그리드에 대해 class의 종류마다 확률 값이 출력되어 이를 토대로 bounding box를 계산한다. 본 연구에서는 [그림2-5]와 같이 추론 시간 대비 평균 정확도를 고려하여 YOLO에서 여러 차례 성능 개선을 거친 YOLO v3 네트워크[17]를 사용하였다.



[그림 2-3] YOLO 모델



[그림 2-4] YOLO의 네트워크 구조



[그림 2-5] 추론 시간과 mAP(IOU@0.5)

제 2 절 가상 키보드 관련 연구

앞서 언급했듯 사용자의 움직임의 제한이 없고 추가비용이 발생하지 않는다는 장점을 가지고 있어 다수의 연구자들이 비전 인식 기반의 알고리즘을 채용한 손동작 인식 기반 가상 키보드 입력 시스템을 제안하였다. Tran et al.[18] Microsoft 사의 Kinetic RGB-D 카메라를 이용하여 얻은 깊이 이미지로부터 손을 추출하여 3DCNN을 통해 손동작을 분류해 가상 키보드의 문자를 입력하는 방식을 제안했으나, RGB-D 카메라가 채용된 전자 장치가 흔치 않아 추가적인 장치 구비가 필요하며, Depth 카메라의 경우 적정수준의 거리가 확보되지 않으면 노이즈가 발생한다는 단점이 있다. Zhang et al.[19]은 RGB 이미지의 손색상 정보를 이용하여 binary 이미지로 만든 뒤 손끝을 기하학적 구조에 따라 추정하고 손끝이 가리키는 가상 키보드의 문자를 출력하였다. 이는 일반적으로 전자기기에 내장된 RGB 카메라 기반 환경에서 사용될 수 있다는 장점이 있지만 색상 정보 기반의 손인식은 배경 색상, 장애물 등의 주변 환경에 의해 오인식이 발생할 수 있다는 단점이 있다.

Lee et al.[20]은 객체 검출 딥러닝 네트워크를 적용해 손의 배경색과 상관없이 손의 ROI를 출력, 손색상 정보를 이용하여 contour 기반의 손동작 인식 방식을 채용했다. 저자는 알파벳의 사용 빈도를 분석하여 3x4 multi-layer의 키보드 layout을 최적화하여 가상 키보드 입력 시간을 줄였으며, [그림 2-6]과 같이 엄지와 검지가 떨어져 있을 때는 Ready 동작, 엄지와 검지가 접촉하였을 때의 손동작을 Click 동작으로 구분했다. 더 나아가 multi-layer 키보드를 효율적으로 제어할 수 있는 multi-layer 손동작을 [그림 2-7]과 같이 제안했다. 각 손동작 클래스 간 구별이 명확하고 직관적이기에 높은 인식률 및 라 입력 속도에 기여했다.

본 논문에서는 Lee et al.[20]에서 제안한 키보드 레이아웃과 손동작 클래스의 정의를 기반으로, 입력 속도 개선을 위한 5개의 모음 손동작 클래스를 추가로 정의하여 가상 키보드를 구현하였다.



(a) Ready 손동작 (b) Click 손동작

[그림 2-6] 클릭 손동작의 과정



(a) Layer1 (b) Layer2 (c) Layer3

[그림 2-7] Multi-layer 손동작

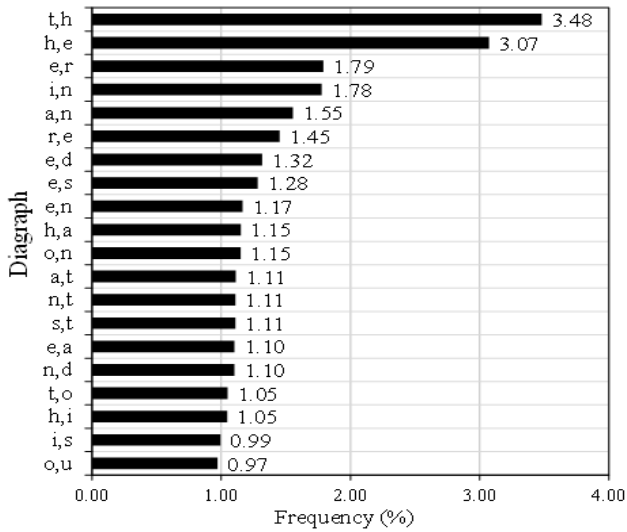
제 3 장 제안한 가상 키보드 시스템

제 1 절 모음 손동작 추가 배경

알파벳의 사용 빈도와 연속적으로 나타나는 두 문자 간의 상관 빈도 통계적 분석을 진행했으며, [표 3-1]과 [그림 3-1]은 일상 생활에서 많이 사용하고 있는 영어 문장 및 e-book을 기준으로 알파벳 사용 빈도 및 상관 빈도를 통계 분석한 내용이다. 모음의 사용 빈도는 평균 39%로 총 26개의 알파벳 가운데 모음은 5개인 점으로 미루어 보아 모음의 사용 비중이 높음을 알 수 있다. 더불어 모음 다음에 자음 또는 자음 다음에 모음이 번갈아 나오는 상관 빈도는 약 65%의 비중을 차지했다. 이러한 특징을 감안하여 각 모음별로 특정된 손동작을 이용해 모음을 입력할 수 있도록 하여 가상 키보드의 입력 시간을 줄이는 방식을 제안한다.

[표 3-1] 모음과 자음의 사용 빈도수

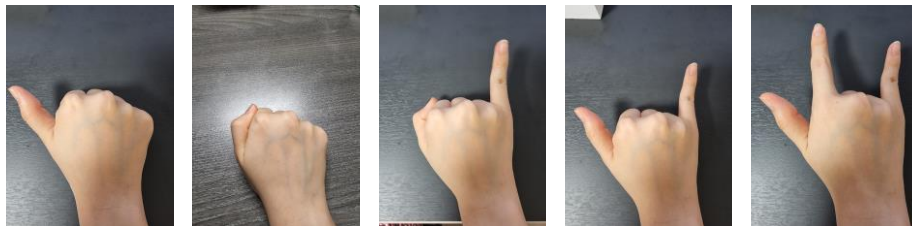
	Vowel	Consonant
500 English [13]	0.41	0.59
Alice [14]	0.39	0.61
20,000 leagues[15]	0.39	0.61
Gulliver [16]	0.38	0.62
LesMiserables[17]	0.38	0.62
average	0.39	0.61



[그림 3-1] 알파벳 상관 빈도수

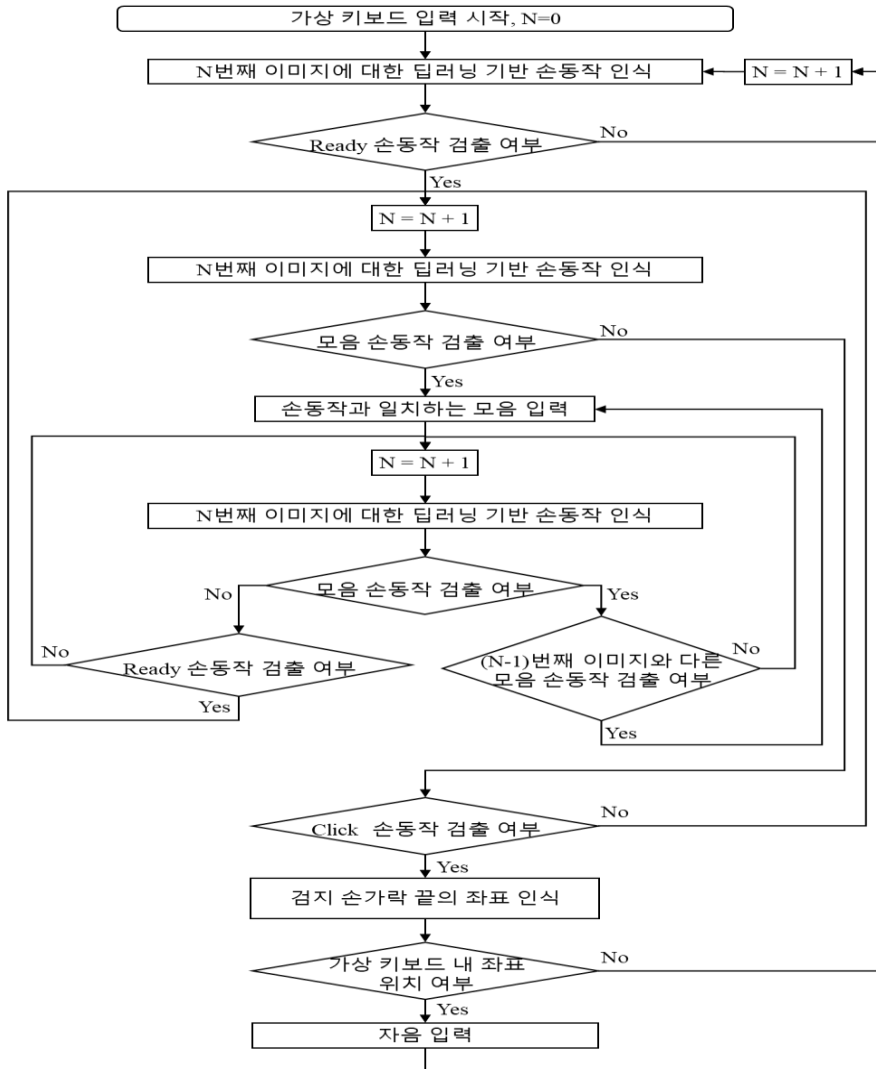
제 2 절 제안한 가상 키보드 입력 시스템

기존의 가상 키보드 입력 방식은 Ready 손동작과 Click 손동작을 인식하여 Click 손동작 검출 시 검지 손 끝이 가리키는 키의 알파벳을 출력한다. 본 논문에서는 기존 방식에 다섯 개의 모음(‘A’, ‘E’, ‘I’, ‘O’, ‘U’)을 입력할 수 있는 모음 손동작을 [그림 3-2]와 같이 추가하여 가상 키보드 시스템을 구현한다. 모음을 입력할 경우 기존의 방식과 달리 검지 손 끝의 이동 시간을 고려하지 않아도 되므로 입력 시간을 줄일 수 있다. 제안한 가상 키보드 입력 시스템의 프레임워크는 [그림 3-3]과 같다.



(a)A 동작 (b)E 동작 (c)I 동작 (d)O 동작 (e)U 동작

[그림 3-2] 모음 손동작



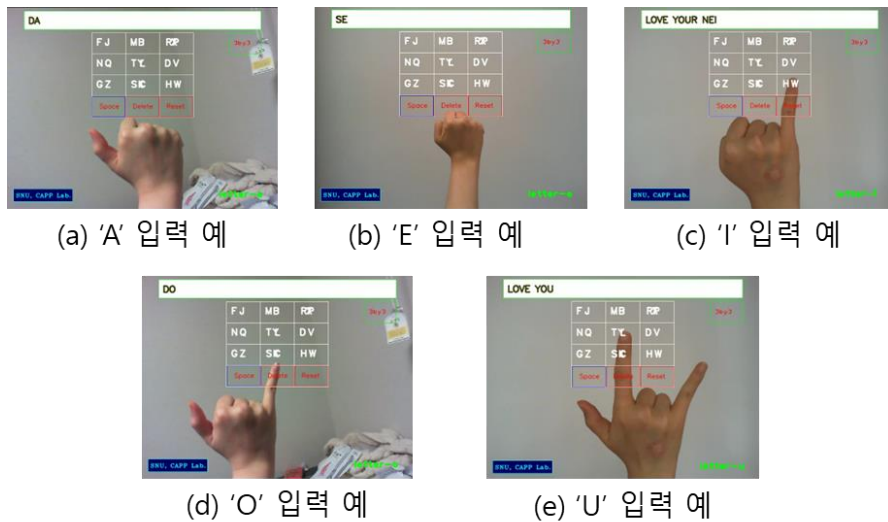
[그림 3-3] 제안한 입력 시스템의 손인식 흐름도

우선 CNN으로 구성된 YOLOv3를 이용하여 손동작 클래스 분류를 진행한다. 본 논문에서는 Lee et al.[20]에서 제안한 손동작에 5개의 모음 손동작을 추가하여 손동작 클래스 분류를 진행한다. 모음 손동작 추가에 따라 분류해야 할 손동작이 늘어남에 따라 정확한 손동작 인식을 위해 수동 및 자동 라벨링을 통한 다양한 환경에서의 손동작 학습 데이터 확보하여 인식율을 향상시키고자 했다. 손동작 인식을 위해 총 133,108장의 11개 클래스의 학습 데이터를 수동 및 자동 라벨링 방식을 통해 생성했다. 학습된 객체 인식 딥러닝 네트워크로 손동작 인식을 진행했다.

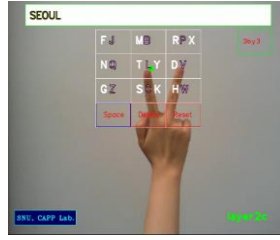
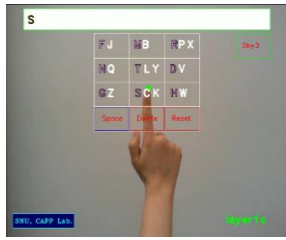
Ready 손동작이 검출된 다음 모음 손동작 또는 Click 손동작이 검출되어야 가상 키보드를 이용한 글자 입력이 가능하도록 하여 키보드 입력을 하지 않을 때의 손동작 인식을 방지, 손동작 오인식에 따른 오타 발생 가능성을 감소하고 딥러닝 인식 오류 또는 손동작 변경 과정에서의 오인식을 방지하고자 하였다.

Ready 손동작 이후 모음 손동작이 검출되면 손동작과 일치한 모음이 입력된다. 이후 새로운 문자를 입력하기 위해서는 Ready 손동작이 검출되어야 하지만, 연달아 모음을 입력하는 경우 서로 다른 모음을 입력하는 경우에 한하여 Ready 손동작 인식이 선행되지 않더라도 입력 가능하도록 설계하여 모음 입력을 효율화 하였다. [그림 3-4]는 가상 키보드 상 모음 입력 예시이다.

Ready 손동작 이후 Click 손동작이 검출되면 multi-layer 손동작에 따라 layer를 선택, 검지 손끝이 가리키는 좌표 정보를 참조하여 자음이 입력된다. [그림 3-5]는 가상 키보드 상 자음 입력 예시이다.



[그림 3-4] 가상 키보드 모음 입력 예시



(a) layer1-'S' 입력 (b) layer2-'L' 입력
 [그림 3-5] 가상 키보드 자음 입력 예시

(c) layer3-'Y' 입력

제 4 장 실험 결과 및 분석

제 1 절 입력 시간 예측 시뮬레이션 결과

입력 시간 예측 시뮬레이션 모델은 Intel i7 CPU와, 64GB RAM를 탑재한 PC를 사용하였고 windows 10 환경에서 개발되었다. [표 4-1]은 다양한 키보드 들과의 성능비교를 위한 시뮬레이션 결과를 보여준다. ‘ABC’, ‘Swarm’, ‘Frequency’ 및 ‘QKLP’, Lee et al.[9],[20]에서 제안한 multi-layer 키보드 레이아웃과 layer-aware gesture 방식의 키보드 입력 방식과 비교하여 제안한 모음 입력 방식을 적용했을 때 입력 시간이 얼마나 개선되는지 평가해 보았다.

시뮬레이션 모델은 Fitts' law에 근거하여 제작되었다. Fitts' law는 사용성 분야에서 인간의 행동에 대해 속도와 정확성 간 관계를 설명하는 기본적인 법칙으로, 시작점에서 목표로 하는 지역까지 얼마나 빠르게 닿을 수 있는지 예측하고자 하는 것으로 Paul M. Fitts에 의해 1954년에 발표된 법칙이다.[21] 이는 사람과 컴퓨터의 상호작용에 대해 연구하는 학문인 HCI(Human Computer Interaction) 분야에서 UI(User Interface)의 효율성을 나타내는 지표로 사용되고 있으며, 키보드 입력하기 위한 인터페이스에서 손의 물리적 이동 시간을 예측하는데 사용된다. Fitts' law의 수식은 (1)과 같다.

$$MT = a+b \cdot \log_2(D/W+1) \quad (1)$$

위의 수식에서 D는 키보드의 키 간의 거리이고, W는 키의 폭이다. 계수 a와 b는 키보드나 인터페이스에 따라 달라질 수 있다. (1)의 식을 제안한 키보드에 적합하게 목적함수를 수정하기 위해 수식 (2)를 추가했다. S(k)는 multi-layer 키보드에서 layer를 선택하는 시간을 나타낸다.

$$S(k) = C(k) + F(k) \quad (3)$$

C(k)는 Click 손동작에 걸리는 시간이고, F(k)는 손동작을 변경할 때 소요되는 시간이다. 계수 a와 b는 0.488, 0.105로, C(k)와 F(k)는 각각 0.488sec, 0.370sec로 Lee et al.[20]과 같이 정의했다.

위 수식을 활용해 제작한 시뮬레이션 모델로 입력 속도를 예측한 결과 모음 손동작을 추가한 입력 방식을 적용했을 때, 평균 20.72%의 입력 속도 향상을 달성하였다. 기존의 키보드 중 가장 빠른 입력이 가능한 layer-aware gesture 방식의 경우 모음 손동작 추가 시,

2.01%의 속도 개선이 있었다. Layer-aware gesture 방식의 속도 개선율이 비교적 낮은 이유는 multi-layer 손동작과 알파벳 사용 빈도수를 기반으로 최적화된 레이아웃이 적용되어, Click 손동작의 반복 및 레이어 변경에 따라 소요되는 시간을 줄일 수 있기 때문이다. 또한 [그림 4-1]의 키보드 레이아웃을 분석해보면 첫 번째 레이어에 5개의 모음 가운데 4개의 모음이 배치되어, 모음 손동작 추가에 따른 이점을 가져가기 어려웠다. 이에 입력시간 단축 효과를 높이기 위해 [그림 4-2]와 같이 모음을 제외한 알파벳을 첫 번째 레이어부터 순차적으로 할당하여 입력 속도를 확인해보니 4.63%의 입력 속도 향상을 보였다.

[표 4-1] 가상 키보드 입력 속도 simulation 결과

* 모음을 3rd layer 로 이동

	Previous						Proposed						
	ABC	Swarm	Frequency	QLKP	Multi-layer	Layer-aware	ABC	Swarm	Frequency	QLKP	Multi-layer	Layer-aware	Layer-aware*
500 English [14]	8,096	7,461	5,955	5,855	5,695	4,079	5,493	5,393	4,746	4,818	4,696	3,992	3,904
Alice [15]	160,863	146,936	118,451	116,184	112,022	81,905	112,317	107,727	93,883	94,479	92,431	80,197	77,959
20,000 leagues [16]	639,534	574,458	457,597	447,143	429,769	319,784	443,576	425,208	363,232	366,263	358,432	313,268	305,112
Gulliver [17]	652,641	586,528	467,386	457,792	443,067	329,652	455,438	438,997	374,872	378,670	371,192	322,910	314,266
LesMiserables [18]	3,568,950	3,193,890	2,554,590	2,481,980	2,395,970	1,786,470	2,480,395	2,361,230	2,028,606	2,041,655	2,001,800	1,750,952	1,703,834

Layer1	Layer2	Layer3																											
<table border="1" style="width: 100%;"> <tr><td>O</td><td>A</td><td>R</td></tr> <tr><td>N</td><td>T</td><td>E</td></tr> <tr><td>I</td><td>S</td><td>H</td></tr> </table>	O	A	R	N	T	E	I	S	H	<table border="1" style="width: 100%;"> <tr><td>F</td><td>M</td><td>P</td></tr> <tr><td>U</td><td>L</td><td>D</td></tr> <tr><td>G</td><td>C</td><td>W</td></tr> </table>	F	M	P	U	L	D	G	C	W	<table border="1" style="width: 100%;"> <tr><td>J</td><td>B</td><td>X</td></tr> <tr><td>Q</td><td>Y</td><td>V</td></tr> <tr><td>Z</td><td>K</td><td></td></tr> </table>	J	B	X	Q	Y	V	Z	K	
O	A	R																											
N	T	E																											
I	S	H																											
F	M	P																											
U	L	D																											
G	C	W																											
J	B	X																											
Q	Y	V																											
Z	K																												

[그림 4-1] layer-aware 키보드 레이아웃

Layer1	Layer2	Layer3																											
<table border="1" style="width: 100%;"> <tr><td>F</td><td>M</td><td>R</td></tr> <tr><td>N</td><td>T</td><td>D</td></tr> <tr><td>G</td><td>S</td><td>H</td></tr> </table>	F	M	R	N	T	D	G	S	H	<table border="1" style="width: 100%;"> <tr><td>J</td><td>B</td><td>P</td></tr> <tr><td>Q</td><td>L</td><td>V</td></tr> <tr><td>Z</td><td>C</td><td>W</td></tr> </table>	J	B	P	Q	L	V	Z	C	W	<table border="1" style="width: 100%;"> <tr><td></td><td></td><td>X</td></tr> <tr><td></td><td>Y</td><td></td></tr> <tr><td></td><td>K</td><td></td></tr> </table>			X		Y			K	
F	M	R																											
N	T	D																											
G	S	H																											
J	B	P																											
Q	L	V																											
Z	C	W																											
		X																											
	Y																												
	K																												

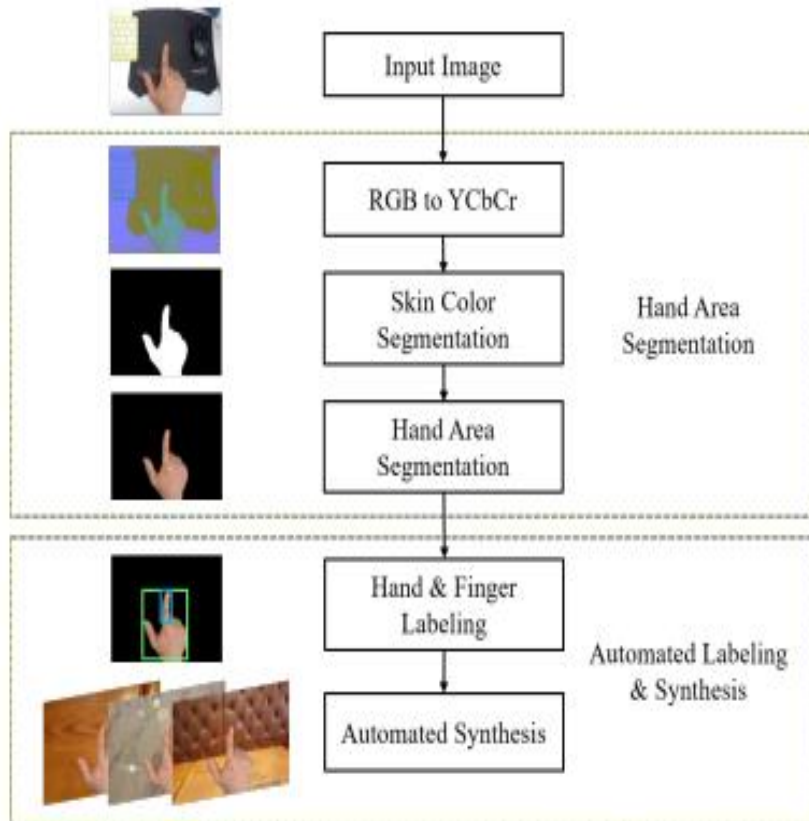
[그림 4-2] 제안한 키보드 레이아웃

제 2 절 딥러닝 네트워크의 학습

제안한 방식은 기존의 가상 키보드 입력에 필요한 손동작에 다섯 개의 모음 손동작의 학습이 필요하다. 손동작이 늘어남에 따른 인식률 저하를 방지하기 위해서는 딥러닝 네트워크 학습 시 다수의 학습 데이터 확보가 필요하다. 또한 배경색, 배경의 복잡도, 손 인식에 방해되는 사물 등 다양한 환경에서의 손동작 인식을 위해서는 학습 데이터의 수뿐만 아니라 손의 크기, 손의 각도, 다양한 배경 등 여러 조건을 포함한 학습 데이터를 확보하는 것이 중요하다.

본 연구에서는 YOLO v3 기반의 손동작 인식을 사용한 가상 키보드를 구현하기 때문에 딥러닝 네트워크 학습에 필요한 데이터 세트를 확보하기 위해서는 데이터 세트를 수동으로 라벨링하는 작업이 필요하다. 앞서 언급한 바와 같이 다수의 다양한 데이터 세트 확보가 필요하기 때문에, 수동 라벨링은 상당한 시간과 노동력이 수반되어야 하는 작업이 된다. 이를 해결하기 위해 실제 손과 다양한 배경을 합성하여 많은 양의 손동작 학습 데이터 세트를 자동 라벨링으로 자체 제작하였다. [그림 4-3]은 손동작 학습 데이터 세트를 생성하고 자동으로 라벨링을 진행하는 과정을 설명한다. RGB 카메라를 통해 각 클래스에 해당하는 손동작을 60초의 동영상으로 녹화한다. 녹화 시 손의 기울기, 원근감 등을 변화시킴으로써 동일한 손동작으로 다양한 조건의 데이터 세트를 얻도록 한다. 녹화된 동영상의 캡처된 RGB 이미지를 YCbCr 이미지로 변환시킨 후 손색상에 해당되지 않는 배경을 제거 손 이미지만 남긴다. 손의 contour에 따라 bounding box를 생성하여 좌표를 얻고 다양한 배경과 손을 합성한다.

자동 라벨링을 통해 12개의 클래스에 부합하는 133,108장의 학습 데이터를 생성했다. 데이터셋은 Ready, Click 손동작 클래스 6개와 검지 클래스, 모음 손동작 클래스 5개로 구성되었으며, 이름은 각각 'layer1s', 'layer1c', 'layer2s', 'layer2c', 'layer3s', 'layer3c', 'index', 'letter-a', 'letter-e', 'letter-i', 'letter-o', 'letter-u'이다. [그림 4-4]는 자동 생성된 손동작 학습 데이터 세트의 예시이다. 학습 데이터 세트를 통해 학습된 딥러닝 네트워크를 자동 라벨링으로 생성한 2,453장의 테스트 데이터셋을 제작하여 검증한 결과 [그림 4-5]와 같이 12개의 클래스에 대한 손동작 평균 인식률은 mAP 94.39% 임을 확인할 수 있었다. 학습 및 검증에 사용된 소프트웨어는 YOLO의 저자가 공개한 오픈 소스 라이브러리인 darknet [17]을 사용했다.



[그림 4-3] 자동 라벨링 흐름도



(a) letter-a



(b) letter-e



(c) letter-i

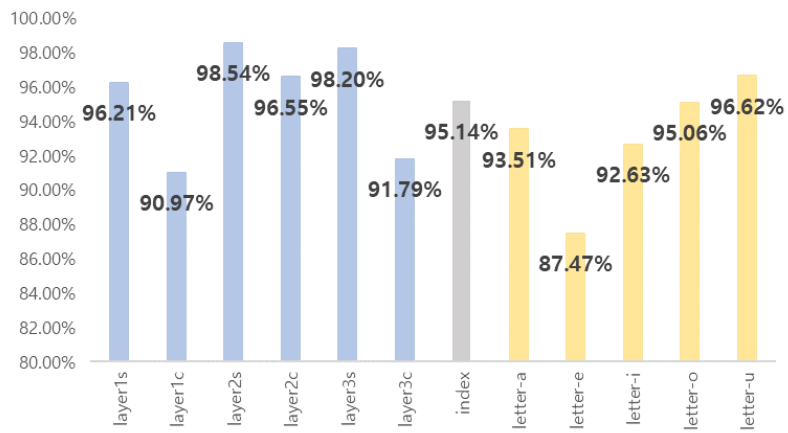


(d) letter-o



(e) letter-u

[그림 4-4] 자동 라벨링으로 생성된 학습 데이터셋 예시



[그림 4-5] 클래스 별 손동작 인식률

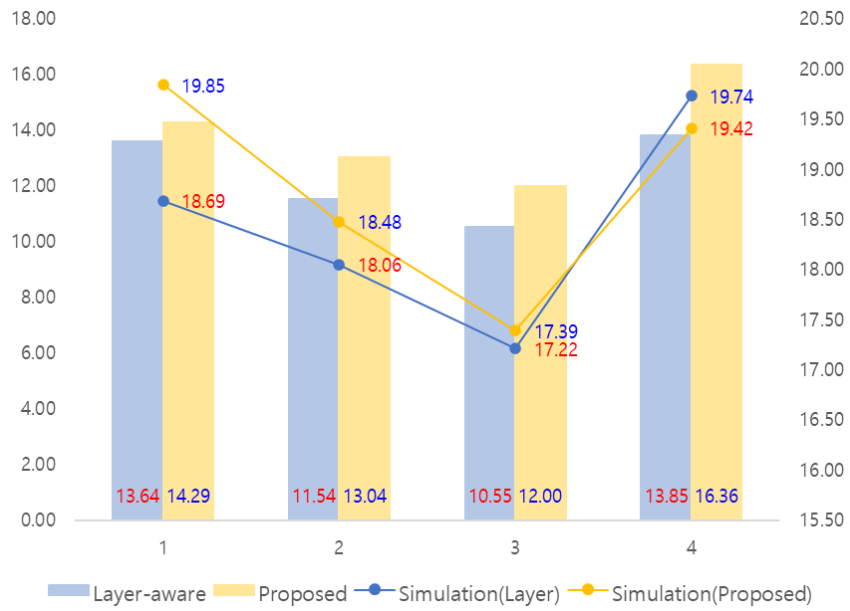
제 3 절 가상 키보드의 사용자 실험 및 분석

앞서 자동 라벨링으로 확보한 데이터 세트를 통해 학습한 객체 인식 딥러닝 네트워크를 사용, 이전에 제안된 layer-aware 입력 방식[20]으로 자음을 입력, 모음 손동작으로 모음을 입력할 수 있도록 하는 알고리즘이 적용된 가상 키보드를 구현하였다. 제안한 가상 키보드 입력 방식의 유효성 검증을 위해 layer-aware 입력 방식과 제안한 모음 손동작 추가 방식의 성능 비교를 진행했다. 각 방식에 대하여 4개의 영어 문장을 10회 타이핑하여 실험을 진행했으며, 실험에 사용된 문장은 [표 4-2]와 같다. 각 문장에는 5개의 모음이 모두 포함되어 있다.

사용자 실험 결과는 [그림 4-6]과 같다. 입력 속도는 WPM(In Words Per Minutes)[22]으로 환산해서 비교했다. 제안한 방식의 입력 속도가 layer-aware 입력 방식 대비 입력 속도를 12.45% 개선함을 확인할 수 있었다. 앞서 1절에서 언급한 시뮬레이션 모델 기반으로의 입력 속도 개선율을 예측했을 때 1.99%로 실제 사용자 실험 시의 개선율이 6배 가량 높음을 확인할 수 있었다. 이는 시뮬레이션 모델의 경우 손가락 이동 거리에 따른 소요시간, layer 변경 및 손동작 변경에 따른 소요 시간만을 고려했으나, 실제 사용자 입력 시 입력이 필요한 알파벳을 배열 내에서 찾는 시간이 추가적으로 소요되기 때문이다. 제안된 키보드 입력 방식의 경우 모음을 배열 내에서 찾지 않아도 되기 때문에 입력 시간이 기대한 바보다 더 큰 폭으로 개선됨을 확인할 수 있었다.

[표 4-2] 사용자 실험 문장

Number	Texts
1	Faith without deeds is useless.
2	Love your neighbor as yourself.
3	Only the just man enjoys peace of mind.
4	A poet is the painter of the soul.



[그림 4-6] 사용자 실험 결과

제 5 장 결론

본 논문에서는 객체 인식 딥러닝 네트워크를 사용하여 손동작을 인식하여 입력 가능한 가상 키보드 입력 시스템을 제안했다. 실생활에서 자주 쓰이는 영어문장과 4권의 e-book 속 문장들을 분석하여 모음의 입력 속도를 향상시켜 가상 키보드 입력 속도를 개선시킬 수 있으리라 기대했다. 이에 모음 입력은 각 모음에 특정된 손동작으로 입력 가능하도록 했다. 시뮬레이션 모델로 기존의 가상 키보드 시스템과 비교 시 평균 20.72% 입력 속도를 개선함을 확인할 수 있었다. 객체 인식 딥러닝 네트워크 기반으로 손동작을 인식함에 있어, 높은 손동작 인식률을 위해 다양한 조건에서의 다량의 학습 데이터 세트를 자동 라벨링을 통해 확보하여 딥러닝 네트워크 학습에 활용했다. 제안한 알고리즘을 적용한 가상 키보드의 사용자 실험을 진행한 결과, 제안하는 방법을 사용하면 기존의 방식보다 12.45% 더 빠른 입력이 가능함을 확인할 수 있었다.

참고 문헌

- [1] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," in *Proc. IEEE ACM Int. Symp. Augmented Reality*, 2000, pp. 111-119.
- [2] T. Lee and T. Hollerer, "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking," in *Proc. 11th IEEE Int. Symp. Wearable Comput.*, 2007, pp. 83-90.
- [3] A. Schreiber and M. Bock, "Visualization and exploration of deep learning networks in 3D and virtual reality," in *Proceedings of HCI International 2019 - Posters*, vol. 1033, 2019, pp. 206-211.
- [4] Y. S. Kim, B. S. Soh, and S. G. Lee, "A new wearable input device: SCURRY," *IEEE Transactions on Industrial Electronics*, Vol. 53, No.6 pp. 1490-1499, 2005.
- [5] A. E. Manjunath, B. P. Vijaya Kumar, and H. Rajesh, "A Preliminary Study of Gestures for Human Computer Interaction(HCI)," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 3, Issue 2, Feb. 2014.
- [6] Ms.Kinjal N. Shah, Mr.Kirit R. Rathod, Mr.Shardul J. Agravat, "A survey on Human Computer Interaction Mechanism Using Finger Tracking," *International Journal of Computer Trends and Technology (IJCTT)*, volume 7, number 3, Jan. 2014.
- [7] Y. T. Hsieh, A. Jylhä, V. Orso, L. Gamberini, G. Jacucci, "Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4203-

4215, May 2016.

- [8] C. Xu and L. Cheng, "Efficient Hand Pose Estimation from a Single Depth Image," *In International Conference on Computer Vision*, pp. 3456–3462, 2013.
- [9] T. Lee and H. Lee, "A New Virtual Keyboard with Finger Gesture Recognition for AR/VR Devices," *in International Conference on Human–Computer Interaction*, pp. 56–67, Jul.2018.
- [10] A. Enkhbat, T. K. Shih, T. Thaipisutikul, N. L. Hakim and W. Aditya, "Handkey: An efficient hand typing recognition using cnn for virtual keyboard", *2020 – 5th International Conference on Information Technology (InCIT)*, pp. 315–319, 2020.
- [11] M. Rahim and J. Shin, "Hand Movement Activity–Based Character Input System on a Virtual Keyboard," *MDPI Journal Electronics*, vol. 9, pp. 774, May. 2020.
- [12] M. Elmezain, A. Al–Hamadi, B. Michaelis, "Spatio–Temporal Feature Extraction–Based Hand Gesture Recognition for Isolated American Sign Language and Arabic Numbers," *IEEE International Symposium on Image and Signal Processing and Analysis*, pp. 254–259, 2009.
- [13] P. Xu, "A real–time hand gesture recognition and human–computer interaction system," *arXiv preprint arXiv: 1704.07296*, 2017.
- [14] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation," *in 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Sep. 2014.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu,

and A. C. Berg, "SSD: Single Shot MultiBox Detector," *in Proceedings of European Conference on Computer Vision*, pp. 21–37, Sep. 2016.

[16] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real–Time Object Detection," *in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[17] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *in arXiv: 1804.02767*, 2018.

[8] D.S. Tran, N. H. Ho, H.J. Yang, E.T. Baek, S.H. Kim and G. Lee, "Real–Time Hand Gesture Spotting and Recognition Using RGB–D Camera and 3D Convolutional Neural Network," *Applied Sciences MDPI*, vol. 10, no. 2, pp. 722, 2020.

[19] Y. Zhang, W. Yan and A. Narayanan, "A virtual keyboard implementation based on finger recognition," *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, 2017.

[20] T. –H. Lee and H. –J. Lee, "Ambidextrous Virtual Keyboard Design with Finger Gesture Recognition," *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, May. 2018

[21] Fitts, P. M., "The information capacity of the human motor system in controlling the amplitude of movements," *J. Exp. Psychol.* 47, 381–391, 1954.

[22] I. S. MacKenzie, "A Note on Calculating Text Entry Speed," [On–line]. Available: <https://www.yorku.ca/mack/RNTextEntrySpeed.html> (Accessed on 25 Feb. 2022)

Abstract

Fast Virtual Keyboard Typing Using Vowel Hand Gesture Recognition

Sujin PARK

Department of Electrical and Computer Engineering
The Graduate School
Seoul National University

This paper proposes a fast virtual keyboard typing method that improves typing method that improves typing speed using hand gesture recognition. In the proposed method, five frequently used English vowels can be quickly input with five dedicated gestures. The proposed method reduces, not only the layer switching time of the multilayer keyboard layout, but also vowel typing time. To predict the performance of the proposed method, simulations are performed considering four scenarios according to the order of appearance of vowels and consonants. The result shows that the proposed method improves typing speed by 20.72%. Lastly user test with typing 4 sample sentences results that typing speed is improved by 12.45% based on WPM.

Keywords : Virtual keyboard, VR, AR, Hand recognition, Object detection, HCI

Student Number : 2021-20860