

노인을 위한 손가락 궤적 기반 제스처 인식 시스템

Finger trajectory-based gesture recognition system for the elderly and the infirm

권성호

동국대학교 컴퓨터공학과

Soungho Kwon

Department of Computer Science and Engineering, Dongguk University

기술의 고도화에 따라 IT 기기의 인터페이스는 점점 더 복잡해져 노인에 대한 사용성이 떨어지고 있다. 이를 극복하고자 NUI의 일종인 제스처를 활용해 인터페이스를 구축하려는 선행 연구는 많았지만 여전히 노인의 특성을 반영하는 것에는 부족함이 있었다. 또한 기존 제스처 인식과 관련한 연구에서 제스처의 시작과 끝에 대한 확인은 어려운 부분이었다. 본 연구는 노인의 특성과 관련한 선행 연구 분석을 통해 스마트 미러에서 신문 기사를 읽는 상황을 가정하고 손가락 궤적 기반 노인을 위한 제스처를 디자인한다. 그 후, 손을 감지할 수 있는 yolo 모델과 손가락 추출과 관련한 선행 연구를 활용하여 영상으로부터 손가락 궤적 추적 및 제스처의 시작과 끝을 인식할 수 있는 시스템을 제안한다. 또한 추출된 손가락 궤적 좌표를 제스처로 사상하는 알고리즘을 제안한다. 다양한 환경에서 제안하는 시스템 설계와 제스처 사상 알고리즘에 대한 검증 실험을 통해 영상 속 배경이 복잡하고 살색과 비슷할수록 제스처 인식이 잘 안 되는 경향을 발견하였고 이와 관련한 향후 연구 방향을 제시하였다.

Key Words : 노인 / 제스처 인식 / 컴퓨터 비전 / 옴로 v3 / 손가락 추적

Abstract

With the advancement of technology, the interface of IT devices is becoming increasingly complex and less usable for the elderly. To overcome this, there were many prior studies to build an interface using gestures, a kind of NUI (Natural User Interface), but still lacking in reflecting the characteristics of the elderly. In addition, identifying the beginning and end of the gesture was a difficult part of the study related to the recognition of existing gestures. This study assumes the situation of reading newspaper articles in smart mirrors and designs gestures for elderly people based on finger trajectory. Subsequently, a hand-sensing yolo model and a prior study involving finger extraction are used to design a system that can recognize the beginning and end of finger trajectory tracking and gesture from images. It also proposes an algorithm to think of extracted finger trajectory coordinates as gestures. Through verification experiments on system design and gesture thought algorithms proposed in various environments, we found that the more complex and similar the background in the images to be to skin, the less likely the gesture recognition was to be, and the future direction of the study was presented.

Key Words : Older / Gesture Recognition / Computer Vision / Yolo v3 / Finger tracking

1. 서론

인구 고령화가 빠르게 진행되면서 노인은 미래의 강력한 소비 주체로 급부상하였다. 통계청의 2013년 총 인구 조사에서 65세 이상 노인의 비율은 12.2%로 지속적으로 증가하여 2050년에는 37.4%에 이를 것으로 전망하고 있다.[1] 이런 추세와는 반대로 기술의 고도화에 따라 IT기기의 인터페이스는 점점 더 복잡해져 노인에 대한 사용성이 떨어지고 있다. 이를 극복하고자 NUI(Natural User Interface)의 일종인 제스처를 활용해 IT기기의 인터페이스를


디자인한 선행연구는 많이 있지만 노인의 특성을 반영하는 것에는 부족함이 있었다. 또한 노인을 위해 디자인 된 제스처를 실질적으로 인식할 수 있는 수준의 시스템 구현에 대한 연구는 더더욱 부족한 실정이다. 특히, 노인의 손 떨림이나 제스처의 시작과 끝의 확인은 기존 제스처 인식과 관련한 연구에서도 어려운 부분이었다. 따라서 본 연구는 노인의 특성과 관련한 선행 연구 분석을 통해 스마트 미러에서 신문 기사를 읽는 상황을 가정하고 손가락 궤적 기반 노인을 위한 제스처를 디자인한다. 그 후, 손을 감지할 수 있는 yolo 모델과 손가락 추출과 관련한 선행 연구를 활용하여 영상으로부터 손가락 궤적 추적 및 제스처의 시작과 끝을 인식할 수 있는 시스템을 제안한다. 또한 추출된 손가락 궤적 좌표를 제스처로 사상하는 알고리즘을 제안한다. 끝으로, 다양한 환경에서 제안하는 시스템 설계와 제스처 사상 알고리즘에 대한 검증 실험을 진행하고 실험 결과 분석 및 향후 연구 방향을 제시한다.

2. 선행연구 분석을 통한 본 연구의 제스처 도출

NUI는 기존의 CUI나 GUI 환경에서 전문 분야의 사람들이나 특별히 학습한 사람들만이 IT 기기를 사용할 수 있었던 것과 달리 특별한 학습과정 없이 손짓, 몸짓, 말소리만으로도 IT 기기를 사용할 수 있게 하여 ‘디지털 신분 장벽을 허무는 21세기 훈민정음’이라고까지 표현되며 사용자 층 확산을 가져올 것으로 기대되었다.[2] 특히, 컴퓨터 환경에 익숙하지 않은 노인이나 신체적 작동에 어려움이 있는 장애인의 독립적인 생활을 위한 기술로서 중요한 역할을 할 것으로 기대된다.[3] 이러한 맥락에서 스마트 TV에서도 NUI의 일종인 제스처를 사용하면 노인 입장에서 더 직관적인 인터페이스를 구축할 수 있을 것으로 기대된다.

한편 노인의 신체적 특성을 분석한 선행 연구에 의하면, 사람은 노화가 진행될수록 동작의 움직임이 느려져 반응시간이 느려지는 특성을 보인다. 또한, 연령 증가에 따라 관절의 활동 범위는 축소되어 좌우 양방향, 전후 방향, 위쪽 방향 등의 먼 쪽에는 손닿기가 어려워진다.[1] 노인과 젊은 세대의 심리적 특성을 비교 분석한 선행 연구를 보면, 젊은 사람들은 창의적이고 추상적 표현을, 노인은 관용적 표현을 보다 익숙하게 여기고 선호하는 경향을 보였다. 노인과 젊은 세대 모두 직관적 표현은 선호하는 것으로 나타났다.[4] 이러한 결과를 종합하면 신체적으로는 활동 범위가 작고 단순한 동작의 제스처를 디자인해야 하고 심리적으로는 관용적 표현, 직관적 표현이 중점이 되도록 제스처를 디자인해야 한다.

본 논문에서는 액티브 시니어를 고려한 스마트 TV 제스처 인터랙션 디자인에 관한 연구에서 도출한 노인 제스처를 참고하여 손가락을 이용한 제스처 중 신체적으로 동작이 편하며 관용적이고 직관적인 표현 동작을 디자인해 이를 인식하는 시스템을 개발한다. 아래 그림은 액티브 시니어를 고려한 스마트 TV 제스처 인터랙션 디자인에 관한 연구에서 도출한 노인 제스처 항목이다.

조작 기능	제스처	조작 기능	제스처	조작 기능	제스처	조작 기능	제스처
①전원 켜기		②전원 끄기		⑤화면 요소 선택		⑥메뉴	
	박수 치기		박수 치기		누르기		손바닥 보여주기
③문자 입력		④취소		⑦홈		⑧볼륨 조절 +	
	타이핑하기		한 손가락으로 X로 하기		삼각형 그리기		한 손을 올리기



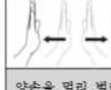


㉔볼륨 조절 -		㉕음 소거		㉖채널 변경 -		㉗화면 확대	
	한 손을 내리기		손을 입이 갖다 대기		한 손을 우측에서 좌측으로 움직이기		양손을 멀리 벌리기
㉘스크롤		㉙채널 변경 +		㉚화면 축소		-	-
	손가락을 위에서 아래로 움직이기		한 손을 좌측에서 우측으로 움직이기		양손을 가까이하기		-

그림 1. 선행연구에서 도출한 스마트 TV 인터페이스 제스처 항목

위 15가지 동작을 참고하여 본 논문에서 구현할 제스처를 아래 표 1과 같이 선정하였다. 본 논문에서 선정한 제스처는 스마트 미러를 통해 신문 기사를 읽는 상황을 가정하여 필요한 동작을 스크롤 위로 이동, 스크롤 아래로 이동, 다음 장으로 넘기기, 이전 장으로 넘기기 4가지로 정하고 각 동작에 대한 제스처를 선정하였다. 선정된 모든 제스처는 물체를 가리킬 때 관용적이고 직관적으로 사용되는 손가락을 사용하여 디자인 하였다. 또한 손 떨림이 있거나 움직임이 느리고 작더라도 인식될 수 있도록 하여 노인으로 하여금 신체적 부담을 최소화 하도록 시스템을 설계하였다.

표 1. 본 논문에서 구현할 제스처

제스처	손가락을 위로 움직이기	손가락을 아래로 움직이기	손가락을 왼쪽으로 움직이기	손가락을 오른쪽으로 움직이기
의미	스크롤 위로 이동	스크롤 아래로 이동	다음 장으로 넘기기	이전 장으로 넘기기

3. 제안하는 손가락 궤적 인식 시스템

1) 전체 시스템 설계

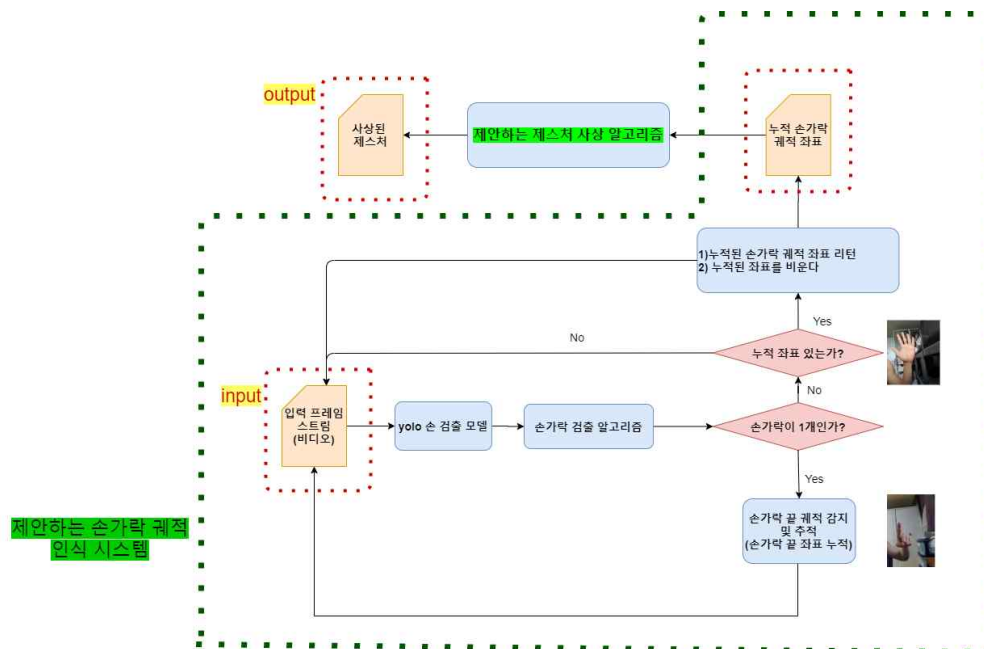


그림 2. 전체 시스템 설계

그림 3의 전체 시스템 설계에서 초록색 점선으로 표시한 부분이 제안하는 손가락 궤적 인식 시스템이다. 입력으로 들어오는 비디오에 대해 각 프레임별로 yolo(you only look once)[5]를 이용해 손 영역을 검출한다. 그 후, 검출된 손 영역에 대해서만 손가락 끝부분을 검출한다. 전체 프레임에 대해 손가락 끝부분을 검출하지 않고 yolo를 통해 먼저 손 영역을 검출한 후 검출된 손 영역에 대해서만 손가락 영역을 추출함으로써, 노이즈를 줄이고 정확도를 높일 수 있었다. 또한 사용자 입장에서 제스처를 취하기 전에 손을 다 편 상태로 카메라에 일정 시간 보여주고 그 후 손가락을 하나만 펴서 제스처를 취한 후 다 끝나면 손을 다시 한 번 펴 줌으로써 제스처의 시작과 끝을 인식할 수 있도록 설계하였다.[6] 위 과정을 통해 인식된 손가락 궤적은 추후 언급될 제안하는 제스처 사상 알고리즘을 통해 특정 제스처로 사상되도록 하였다. 설계의 각 요소별 상세내용은 뒤에서 자세히 다룬다.

2) yolo 손 검출 모델

yolo는 실시간 영상에서 객체감지를 위해 고안된 딥러닝 모델이다. 모델 구조는 그림 3에서 볼 수 있듯이 입력 비디오를 프레임 단위로 읽어 각 프레임마다 yolo 모델이 감지한 객체에 대해 bounding box, score, label을 반환하는 구조이다.

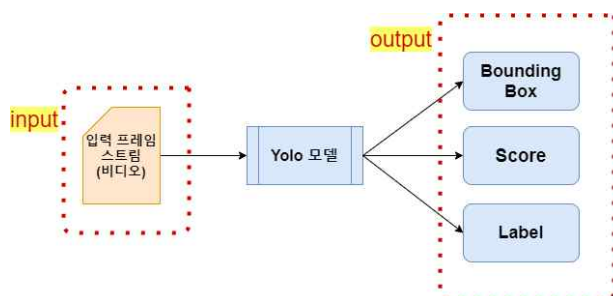


그림 3. yolo 모델 구조

여기서 label은 감지된 객체의 클래스 이름이고 bounding box는 감지된 객체의 위치를 해당 프레임에 대한 좌표로 표현한 것이다. 마지막으로 score는 감지된 객체에 대한 신뢰도를 0~1사이의 비율로 표현한 것이다. 본 연구에서 yolo가 감지해야 할 객체는 손 하나뿐이다. 즉, 그림 3에서 yolo 모델이 반환하는 label의 종류는 손 하나이다.

손을 인식하는 모델은 <https://github.com/cansik/yolo-hand-detection>[7](2020-07-29)를 참고하여 이미 학습된 웨이트를 변환하여 사용하였다. 사용한 모델에 대한 성능 지표는 그림 4와 같다.

Precision: 0.89 Recall: 0.85 F1-Score: 0.87 IoU: 69.8

그림 4. 본 연구에서 사용한 yolo모델 성능 지표

본 연구에서는 yolo모델이 감지한 손에 대한 score값이 0.2 이상인 경우에만 해당 bounding box를 손으로 인식하여 다음에 소개할 손가락 검출 알고리즘을 적용하도록 설계하였다.

3) 손가락 검출 알고리즘

앞선 장에서 yolo를 통한 손 영역이 잘 검출되었다면 그 영역에 대해서만 손가락 검출 알

고리즘을 적용한다.

본 논문의 손가락 검출 알고리즘은 <https://webnautes.tistory.com/1378>[8](2020-07-30)을 참고하였다.

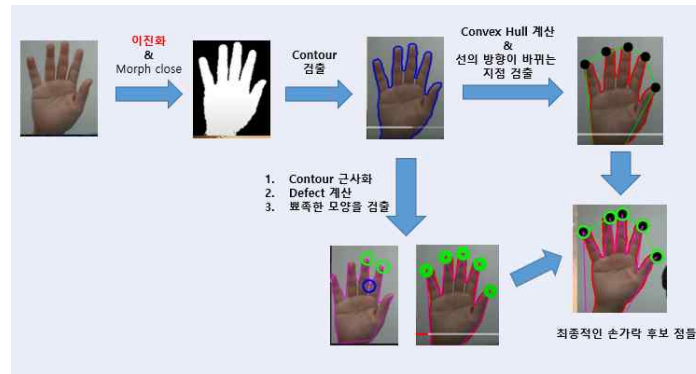


그림 5. 손가락 검출 알고리즘

yolo를 통해 검출된 손 영역에 대해 OpenCv가 제공하는 배경 제거 알고리즘을 사용해 손 영역만 흰색으로 분류한다. 다음으로 흰 부분의 외곽선을 따라 Contour를 검출하고 가장 큰 Contour만 추출한다. 추출된 Contour에서 Convex Hull을 계산해 선의 방향이 바뀌는 지점을 손가락 후보 영역으로 검출한다. 이 방법만으로도 손가락이 잘 검출되는 듯 보이나 손가락을 굽힐 시 손가락이 아닌 부분도 인식되는 오류가 발생한다.



그림 6. Convex Hull 계산을 통한 손가락을 추출한 경우 발생하는 오류 샘플

한편, Contour를 근사 시키고 defect를 계산한 후 두 개의 손가락 사이에는 반드시 defect가 존재한다는 특성을 사용해 뾰족한 모양을 손가락 후보로 선정할 수도 있다. 아래 그림에서 파란 부분은 defect, 초록 부분은 해당 defect로 인해 결정된 손가락 후보이다. 파란 부분의 각도가 90도 이하일 경우에만 초록 부분을 손가락 후보로 결정한다. 하지만 이 방법은 손가락을 하나만 펼 경우 검출하지 못한다.

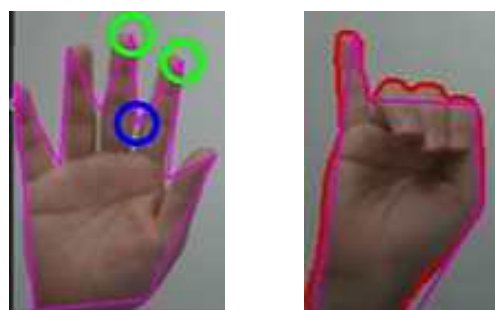


그림 7. defect 계산을 통한 손가락 후보 결정(왼쪽)과 발생하는 오류(오른쪽) 샘플

따라서 Convex Hull로 구한 손가락 후보와 defect로 구한 손가락 후보의 합집합을 최종 손가락 후보로 정한다. 이렇게 하면 모든 손가락을 검출할 수 있지만 아직 제거해야 할 손가락 후보가 존재한다. 최종 손가락 후보 중 양옆의 가장자리 각도가 90도 이하인 손가락 후보만 최종적인 손가락 영역으로 확정된다. 아래 그림에서 파란색 부분은 손가락 후보이지만 양옆의 가장자리 각도가 90도 이상이기 때문에 손가락 영역에서 제외된다.

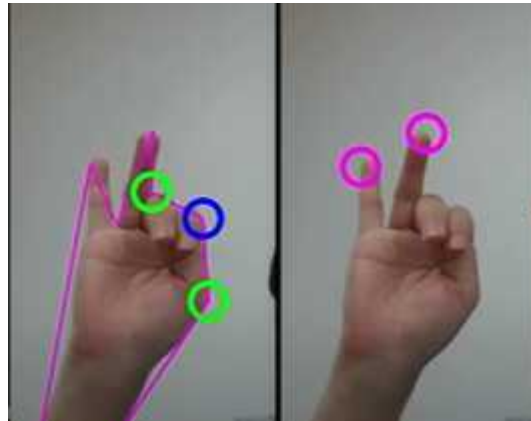


그림 8. defect 계산을 통한 검출 시 양옆 defect의 각도 확인 샘플

4. 제안하는 제스처 사상 알고리즘

위에서 제시한 손가락 검출 방법과 제안하는 손가락 궤적 인식 시스템을 통해 영상으로부터 프레임 단위로 손가락 궤적을 추적하여 그 위치를 누적할 수 있다. 누적된 손 궤적 좌표를 제스처로 사상 하는 알고리즘은 아래와 같다.

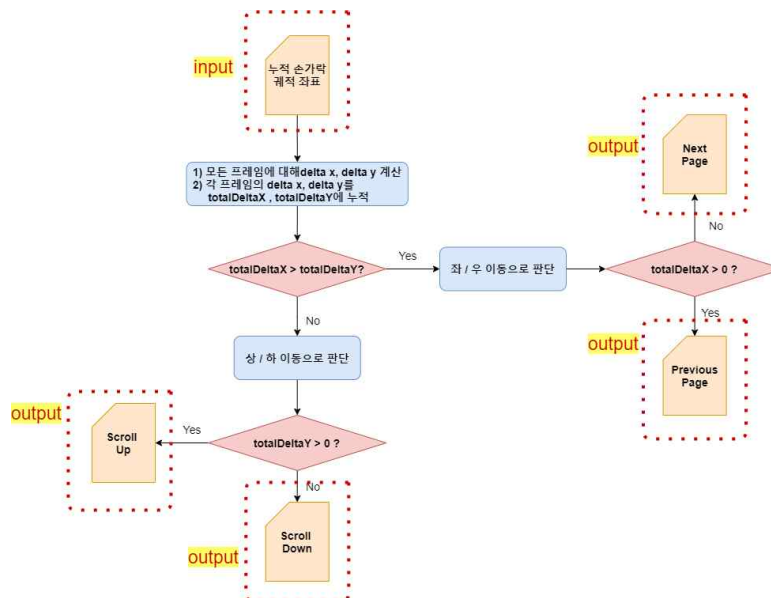


그림 9. 제스처 사상 알고리즘

프레임은 2차원이기 때문에 손가락 위치는 (x좌표, y좌표) 형태로 표현된다. 누적된 손 궤적 좌표에서 각 프레임마다 좌표의 x축 증분과 y축 증분을 구한다. 그 후 각 축에 대한 증

분을 모두 합해 x값의 총 증분과 y값의 총 증분을 구하여 더 큰 증분을 가진 축이 무엇인지 결정한다. 또한 증분의 부호를 통해 최종적으로 상, 하, 좌, 우 이동 방향을 결정하는 알고리즘이다. 해당 알고리즘을 통해 손 떨림이 있거나 작은 움직임을 취한 손 궤적 좌표에 대해서도 정확한 제스처 사상이 가능하도록 하였다.

5. 시스템 검증 실험

1) 실험 환경

스마트 미러에서 신문 기사를 읽는 상황을 가정하여 실험을 진행한다. 따라서 사용자는 카메라와 마주 보고 있고 얼굴과 한 손이 카메라에 나타난다고 가정한다. 이러한 가정 하에 배경 색, 카메라와 손의 거리 2가지 요소의 변화에 따른 제스처 인식의 정확도를 검증하는 실험을 진행한다. 배경 색은 검은색, 파란색, 흰색, 노란색, 복잡한 배경 5가지로 구분한다. 거리는 가까운 거리(카메라로부터 60cm 떨어진 거리), 멀리 떨어진 거리(카메라로부터 160cm 떨어진 거리) 2가지로 구분한다. 실험은 연구자의 집에서 찍은 손가락 끝으로 제스처 동작을 취하는 영상 데이터를 기반으로 진행되었다.



그림 10. 실험의 배경 조건

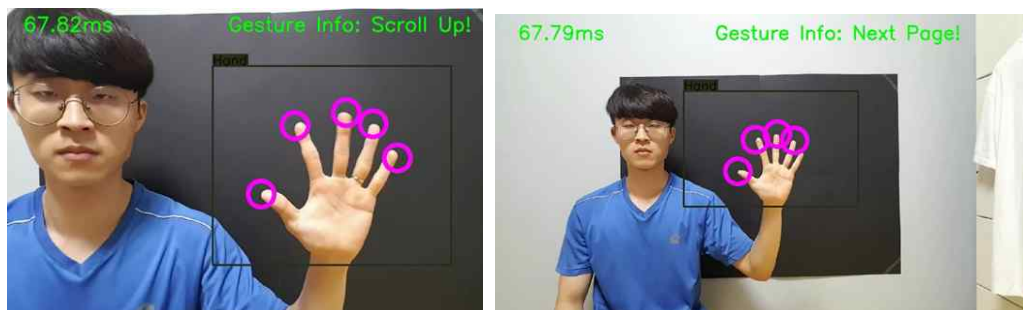


그림 11. 실험의 거리 조건(카메라로부터 60cm 떨어진 거리와 160cm 떨어진 거리)

2) 실험 결과

표 2. 실험 결과

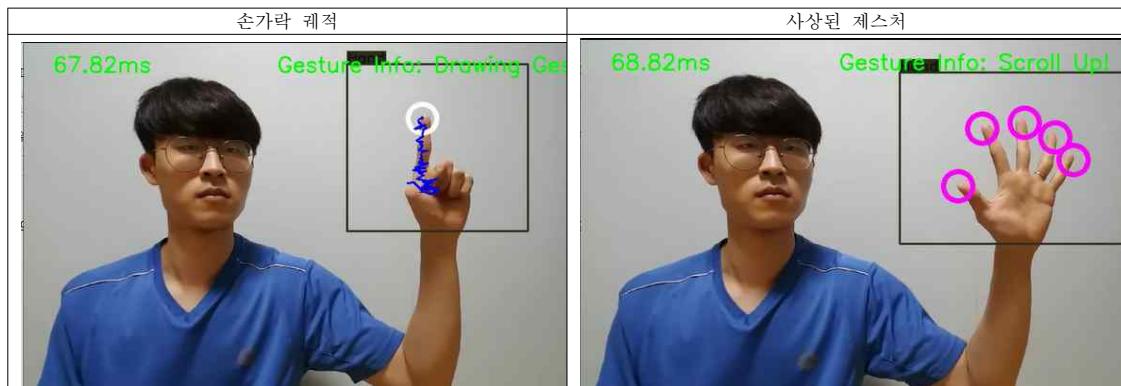
가까운 거리	흰색	검은색	노란색	파란색	복잡한 배경	total
scroll up	10 / 10	10 / 10	6 / 10	10 / 10	6 / 10	42 / 50
scroll down	10 / 10	10 / 10	8 / 10	9 / 10	4 / 10	41 / 50
next page	8 / 10	8 / 10	5 / 10	9 / 10	4 / 10	34 / 50
right page	9 / 10	9 / 10	6 / 10	10 / 10	4 / 10	38 / 50
total	37 / 40	37 / 40	25 / 40	38 / 40	18 / 40	155 / 200

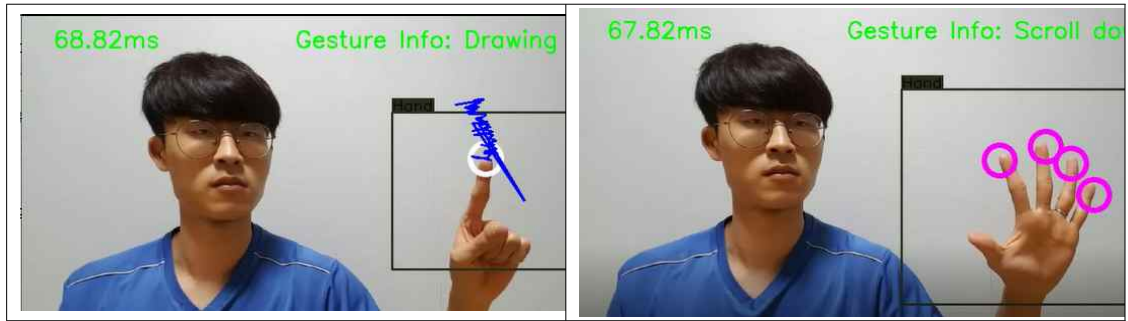
먼 거리	흰색	검은색	노란색	파란색	복잡한 배경	total
scroll up	8 / 10	10 / 10	6 / 10	10 / 10	6 / 10	40 / 50
scroll down	8 / 10	10 / 10	8 / 10	9 / 10	4 / 10	39 / 50
next page	9 / 10	9 / 10	6 / 10	8 / 10	2 / 10	34 / 50
right page	10 / 10	8 / 10	7 / 10	10 / 10	3 / 10	38 / 50
total	35 / 40	37 / 40	27 / 40	37 / 40	15 / 40	151 / 200

3) 실험 결과 분석

실험은 총 400번의 제스처에 대한 인식의 정확도를 테스트하는 것으로 진행되었는데 그중 306개가 정확히 인식되어 총 인식률은 76.5%를 보였다. 가까운 거리일 때 77.5%, 먼 거리일 때 75.5%의 인식률을 보여주었다. 즉, 거리에 따른 인식률의 차이는 거의 존재하지 않았다. 하지만 배경에 따른 인식률은 크게 차이가 나는 경향을 보였다. 검은 배경, 파란 배경, 흰 배경의 경우 90% 이상의 인식률을 보였으나 노란 배경의 경우 65%, 복잡한 배경의 경우 41%의 정확도를 보였다. 정확도가 높게 나타나는 배경 색의 경우 배경이 단조로울 뿐 아니라 살색과 배경색이 대비되어 손이 확실하게 구분되는 특징을 가졌다. 반대로 낮은 정확도를 보인 노란색 배경의 경우 살색과 비슷한 색상이기 때문에 yolo에 의한 손 감지가 잘되지 않아 부정확한 실험 결과를 얻은 것으로 분석된다. 복잡한 배경의 경우 손 감지가 잘되지 않을 뿐 아니라 배경 제거 알고리즘도 정상적으로 작동하지 않아 더 낮은 정확도를 보이는 것으로 분석된다. 결론적으로, 손이 두드러지는 단조로운 배경에서의 인식률은 매우 높지만 배경이 복잡하거나 살색과 비슷할수록 인식률이 낮아지는 경향을 발견하였다. 한편, 아래 표 3에서처럼 손 떨림이 있는 손 궤적을 제스처로 사상할 경우 제안하는 제스처 사상 알고리즘이 매우 잘 동작하는 것을 알 수 있었다.

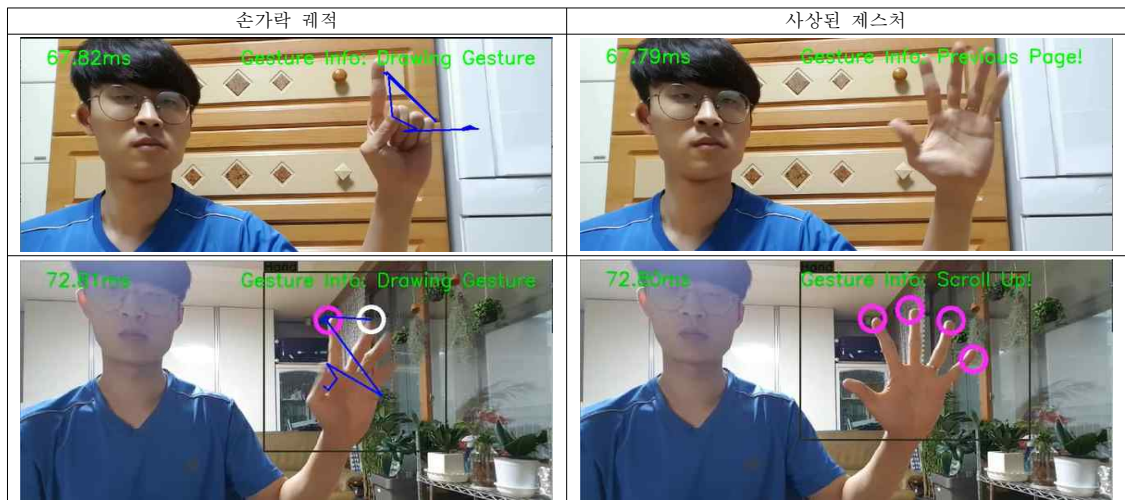
표 3. 손 떨림 손가락 궤적에 대해 사상된 제스처 예시





반대로 표 4는 제스처 사상이 잘 안 되는 경우에 대한 예시이다.

표 4. 제스처 사상이 잘 안 되는 경우에 대한 예시



실험결과에서 언급했듯이 배경색이 손과 비슷하거나 복잡한 배경의 경우 손가락 꺾적에 대한 추적이 잘 안 되는 경향이 있었다. 이로 인해 손가락 꺾적 좌표를 누적하는 과정에서 좌표가 일정한 방향성과 비율로 변화하지 않고 튀는 값이 한 번씩 나오는 것을 확인할 수 있었다. 튀는 값은 각 축에 대한 증분을 계산할 때 결정적인 영향을 줄 가능성이 크다. 결론적으로 손가락에 대한 불안정한 추적은 불안정한 꺾적을 만들게 되고 이로 인해 제스처 사상이 잘 되지 않는 것으로 판단된다.

6. 결론 및 향후 연구

본 연구에서는 스마트 미러에서 뉴스 기사를 보는 상황을 가정하고 필요 기능을 스크롤 위, 스크롤 아래, 이전 페이지로 이동, 다음 페이지로 이동 4가지로 정의 후 각 기능에 대해 노인을 위한 적절한 제스처를 선정하여 이를 인식할 수 있는 시스템을 구현했다. 시스템 구현을 위해 yolo를 사용해 손을 검출 후 손 영역에 대해서만 손가락 검출 알고리즘을 적용하여 손가락을 검출하였고 이를 추적하여 제스처를 인식하게 만들었다. 이에 대한 실험에서, 전체 시스템의 제스처 인식의 정확도는 75%였는데, 배경이 복잡하고 살색과 비슷할수록 정확도가 감소하는 경향을 보였다. 이는 살색 배경에서 yolo의 손 검출의 정확도가 떨어지고 복잡

한 배경일수록 배경제거 알고리즘이 잘 동작하지 않기 때문이었다. 이 부분은 더 좋은 object detection 모델과 양질의 손 데이터 셋을 통한 학습을 통해 더 높은 정확도를 얻을 수 있을 것으로 기대된다.

한편, 각종 IoT 장비나 웨어러블 센서를 동원해 제스처를 인식하는 시스템 개발 또는 라이프 로그[7] 수집 및 분석을 통한 위험 인지 시스템 개발에 대한 연구는 활발히 진행되고 있다. 본 연구 또한 이러한 통합 시스템에 포함되어 복합적인 정보를 주고받을 수 있다면 발전 가능성과 활용도는 더 클 것으로 기대된다. 일례로, 영상 데이터 뿐 아니라 구체적인 손의 위치를 웨어러블 센서를 통해 얻거나 손의 이동 방향 및 속도를 알 수 있다면 더 정밀한 손가락 추적이 가능할 것으로 판단된다. 손가락 추적이 더 정밀해지면 본 연구에서 제시한 제스처 말고도 손가락을 사용한 수많은 제스처를 인식할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] 이현주, 액티브 시니어를 고려한 스마트TV 제스처 인터랙션 디자인에 관한 연구, 국내석사학위논문 경성대학교 디지털디자인전문대학원, 2014.
- [2] 송해원, 비 접촉 인터랙션을 위한 제스처 디자인 가이드 연구, 서울과학기술대학교 NID융합기술대학원 석사학위 논문, 2012.
- [3] 이상미, “복수의 가전제품에서 사용할 수 있는 통합 제스처 UI개발 연구: TV, 에어컨, 오디오의 사용 환경을 중심으로,” HCI 2013 학술대회, 강원도 하이원 리조트, 2013.
- [4] 김현정, “스마트 TV 제스처 인터랙션의 세대별 인지모델 비교 연구,” 한국디자인문화학회지, 22(3), pp. 127-142, 2016.
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unifide, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.
- [6] Sohom Mukherjee, Sk. Arif Ahmed, Debi Prosad Dogra, "Fingertip Detection and Tracking for Recognition of Air-Writing in Videos," Expert Systems with Applications, vol. 136, pp. 217-229, 2018.
- [7] <https://github.com/cansik/yolo-hand-detection>(2020-07-29)
- [8] <https://webnautes.tistory.com/1378>(2020-07-29)
- [9] 임해원, 이현수, "IT 기술을 활용한 독거노인 고독사 관리 서비스에 관한 연구," 한국

실내 디자인학회논문집, 27(3), pp. 71-80, 2018.