



노인은 스마트 TV의 다양한 기능에 대한 인터페이스가 너무 어려워 항상 사용하는 기능, 단순한 기능만 사용하기 쉽다. 따라서 노인 입장에서 보다 이해하기 쉽고 동작하기 편안한 제스처 인터페이스를 구축할 수 있다면 노인의 스마트 TV의 사용성도 증가할 것으로 기대된다.

이러한 추세와 맞물려 최근 스마트 TV 환경에서 노인의 접근성을 고려한 유니버설 제스처 디자인에 대한 연구가 활발히 진행되고 있다.[2] 본 논문은 이러한 선행 연구들로부터 노인의 신체, 심리적 특성을 분석하고 도출된 제스처를 참고하여 스마트 TV 환경에서 손가락을 이용한 제스처 몇 가지를 선정한다. 선정된 제스처를 인식할 수 있는 시스템을 설계 및 구현해보고 다양한 환경에서 실험을 통해 유효성을 검증한다. 마지막으로 실험 결과 분석 및 향후 연구 방향에 대해 논한다.

2. 선행연구 분석을 통한 본 연구의 제스처 도출

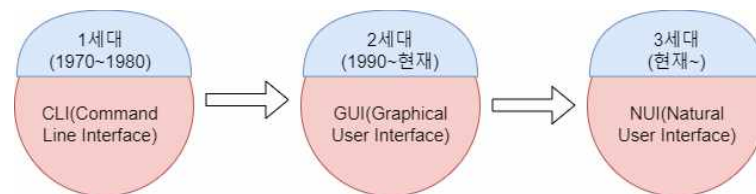


그림 1. UI의 진화 과정

NUI(Natural User Interface)는 기존의 CLI나 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.[3] 특히, 컴퓨터 환경에 익숙하지 않은 노인이나 신체적 작동에 어려움이 있는 장애인의 독립적인 생활을 위한 기술로서 중요한 역할을 할 것으로 기대된다.[4] 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 신체적 특성을 보인다. 또한, 연령 증가에 따라 관절의 활동 범위는 축소되어 좌우 양방향, 전후 방향, 위쪽 방향 등의 먼 쪽에는 손닿기가 어려워진다.[1] 또한, 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.[2] 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 선정 및 추가해 이를 인식하는 시스템을 개발한다. 아래 그림은 액티브 시니어를 고려한 스마트 TV 제스처 인터랙션 디자인에 관한 연구에서 도출

한 노인 제스처 항목이다.

조작 기능	제스처	조작 기능	제스처
①전원 켜기	 박수 치기	②전원 끄기	 박수 치기
③화자 입력	 타이핑하기	④취소	 한 손가락으로 X표 하기
⑤화면 요소 선택	 누르기	⑥뒤로	 손바닥 보여주기
⑦홈	 삼각형 그리기	⑧볼륨 조절 +	 한 손을 움직이기
⑨볼륨 조절 -	 한 손을 내리기	⑩슬라이더	 손을 밑에 갖다 대기
⑪스크롤	 손가락을 위에서 아래로 움직이기	⑫화면 회전 +	 한 손을 좌측에서 우측으로 움직이기
⑬화면 회전 -	 한 손을 우측에서 좌측으로 움직이기	⑭로터 회전	 양손을 벌려 펼치기
⑮화면 축소	 양손을 가까이하기	-	-

그림 2. 선행연구에서 도출한 스마트 TV 인터페이스 제스처 항목

위 15가지 동작을 참고하여 본 논문에서 구현할 제스처를 아래 표와 같이 선정하였다. 본 논문에서 선정한 제스처는 스마트 TV를 통해 e-book을 읽는 상황을 가정하여 필요한 동작을 스크롤 위로 이동, 스크롤 아래로 이동, 다음 장으로 넘기기, 이전 장으로 넘기기 4가지로 정하고 각 동작에 대해 제스처를 선정하였다. 선정된 모든 제스처는 작은 손가락 움직임만으로도 인식될 수 있도록 하여 노인으로 하여금 신체적 부담을 최소화 하도록 시스템을 설계할 것이다.

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. 전체 시스템 설계

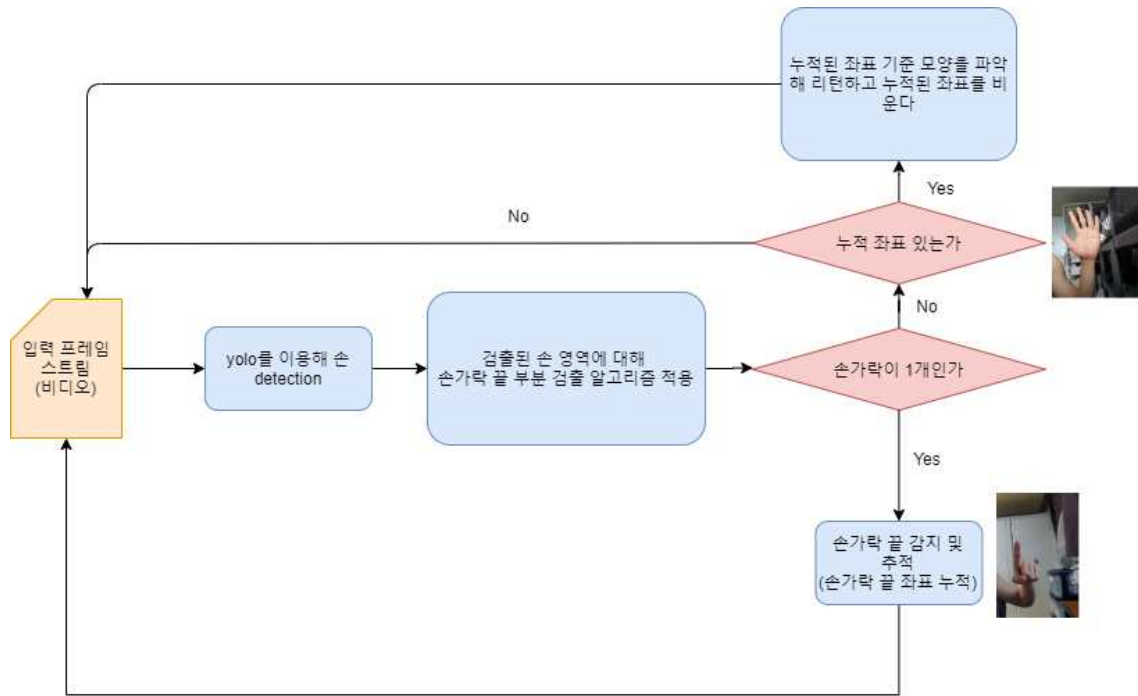


그림 3. 전체 시스템 설계

전체 시스템 설계를 위 그림과 같이 하였다. 입력으로 들어오는 비디오에 대해 각 프레임별로 yolo[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 추출한 후 추출된 손 영역에서만 손가락 영역을 추출함으로써, 노이즈를 줄이고 정확도를 높일 수 있었다. 또한 사용자 입장에서 제스처를 하기 전에 손을 다 편 상태로 카메라에 일정 시간 보여주고 그 후 손가락을 하나만 펴서 제스처를 취한 후 다 끝나면 손을 다시 한번 펴 줌으로써 제스처의 끝을 알려주도록 설계하였다.[6]

4. Yolo 모델 학습

yolo 모델이 인식하고자 하는 객체는 손이다. 손은 영상 속에서 다양한 모습으로 나타날 수 있고 신체에서 크기가 작은 부분이기 때문에 비교적 학습이 어려운 객체이다.



그림 4. 학습이 어려운 손 이미지 샘플

따라서 손이 영상 속에서 잘 나타나는 특정 상황을 가정하여 학습을 진행하였다. 스마트 TV에서 e-book을 읽는 상황을 본 논문에서는 가정했기 때문에 어느 정도 손이 카메라에서 가까이 있다는 가정을 만들었고 이에 대한 데이터를 준비하여 학습시켰다.

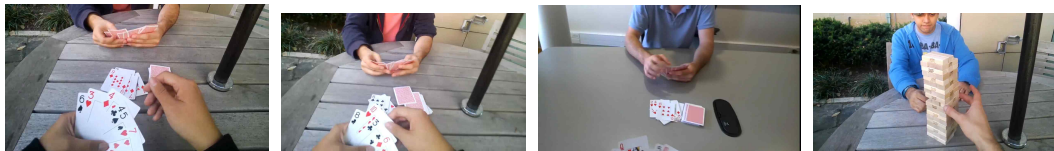


그림 5. 학습에 사용한 손 이미지 샘플

학습에는 <http://vision.soic.indiana.edu/projects/egohands/> (2020-07-22)의 데이터를 사용하였다. 해당 데이터로 batch size를 6으로 하고 300epoch 만큼 학습을 진행한 결과 loss는 0.4, mAP는 0.83인 모델을 얻을 수 있었다. 아래 사진은 학습 결과로 얻은 모델을 테스트한 결과 예시이다.

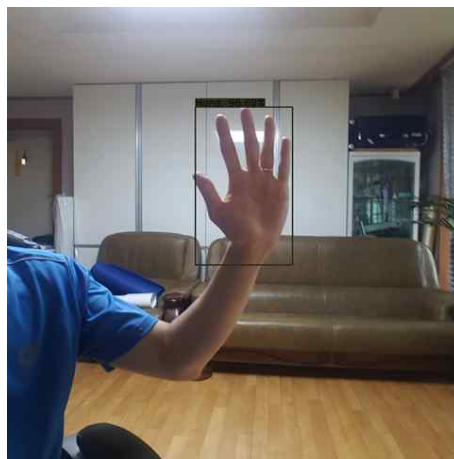


그림 6. yolo 학습 결과 테스트 이미지 샘플

5. 손가락 검출 알고리즘

앞선 장에서 yolo를 통한 손 영역이 잘 검출되었다면 그 영역에 대해서만 손가락 검출 알고리즘을 적용한다.

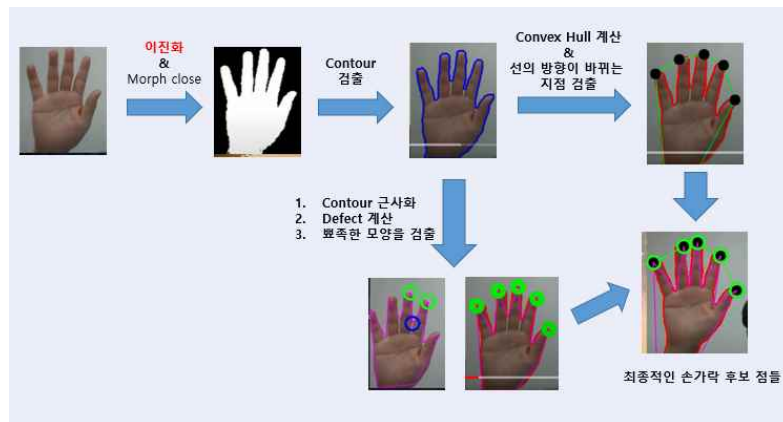


그림 7. 손가락 검출 알고리즘

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 8. 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.

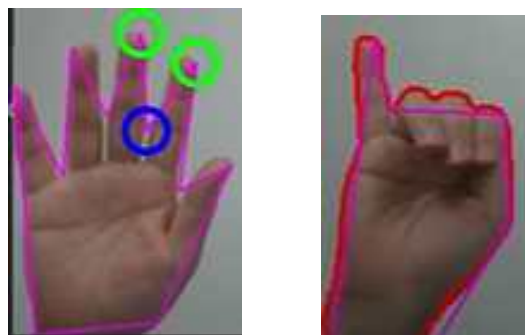


그림 9. defect 계산을 통한 손가락 후보 결정(왼쪽)과 발생하는 오류(오른쪽) 샘플

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.

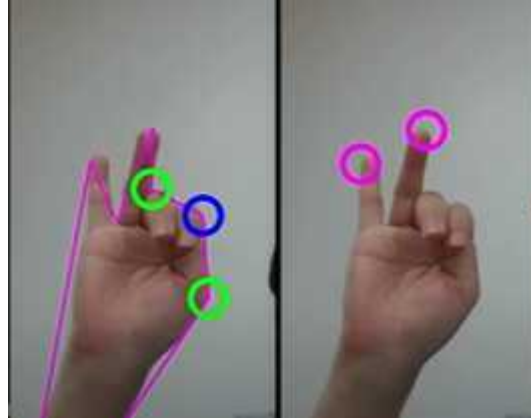


그림 10. defect 계산을 통한 검출 시 양옆 defect의 각도 확인 샘플

5. 제스처 매핑

위에서 제시한 손가락 인식 방법과 전체 시스템 설계를 통해 영상으로부터 프레임 단위로 손가락을 추적하여 그 위치를 누적할 수 있다. **누적된 손가락** 위치를 제스처로 매핑 하는 알고리즘은 아래와 같다.

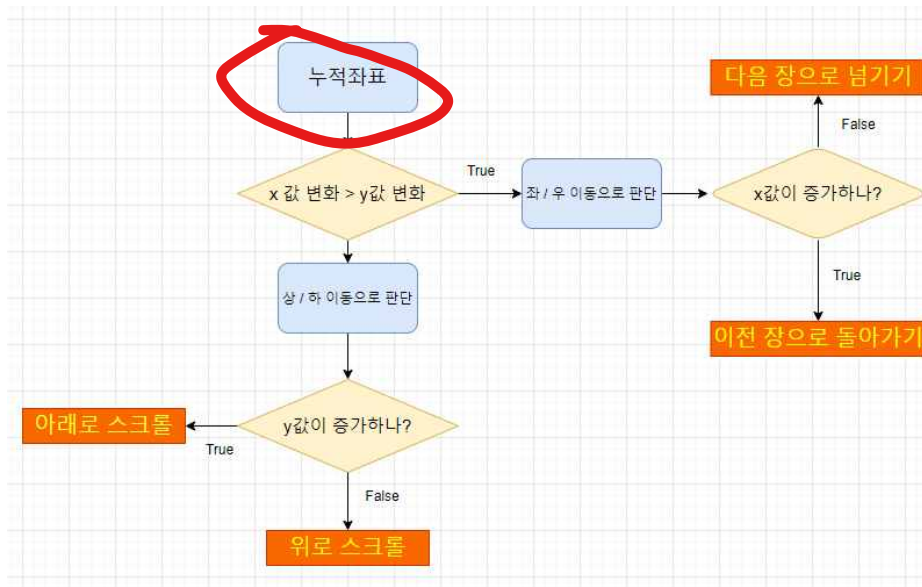


그림 11. 제스처 매핑 알고리즘

프레임은 2차원이기 때문에 손가락 위치는 (x좌표, y좌표) 형태로 표현된다. 누적 좌표에서 각 프레임마다 좌표의 x축 증분과 y축 증분을 구한다. 그 후 각 축에 대한 증분을 모두 합

해 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.

6. 시스템 검증 실험

1) 실험 환경

스마트 TV에서 e-book을 보는 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, 멀리 떨어진 거리 2가지로 구분한다. 실험은 연구자의 집에서 찍은 손가락 끝으로 제스처 동작을 취하는 영상 데이터를 기반으로 진행되었다.



그림 11. 실험의 배경 조건

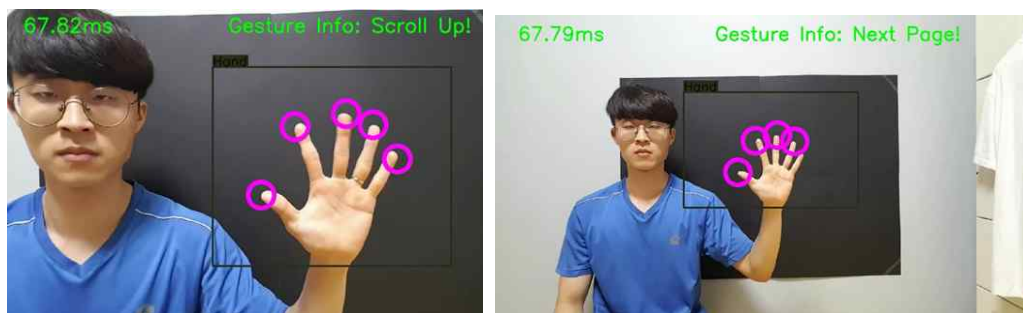


그림 12. 실험의 거리 조건

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.

7. 결론 및 향후 연구

본 연구에서는 스마트 TV에서 e-book을 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 인식 후 손 영역에 대해서만 손가락 추출 알고리즘을 적용하여 손가락을 추출하였고 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 감지의 정확도가 떨어지고 복잡한 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 정밀해 지면 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] 김현정. "스마트 TV 제스처 인터랙션의 세대별 인지모델 비교 연구." 한국디자인문화학회지, 22(3), 127-142, 2016.
- [3] 송해원. "비 접촉 인터랙션을 위한 제스처 디자인 가이드 연구.", 서울과학기술대학교 NID융합기술대학원 석사학위 논문, 2012.
- [4] 이상미. "복수의 가전제품에서 사용할 수 있는 통합 제스처 UI개발 연구: TV, 에어컨, 오디오의 사용 환경을 중심으로.", HCI 2013 학술대회, 2013.
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unifide, Real-Time Object Detection." University of Washington, 2016.
- [6] Sohom Mukherjee, Sk. Arif Ahmed, Debi Prosad Dogra. "Fingertip Detection and Tracking for Recognition of Air-Writing in Videos." National Institute of Technology Durgapur, 2018.
- [7] 임해원, 이현수. "IT 기술을 활용한 독거노인 고독사 관리 서비스에 관한 연구." 한국 실내디자인학회논문집, 27.3: 71-80, 2018.