

ML 기반의 영상처리를 통한 알람 프로그램

*김덕민 *정현우 **박구만

서울과학기술대학교

*forfake@naver.com

Alarm program through image processing based on Machine Learning

*Kim, Deok-Min *Chung, Hyun-Woo **Park, Goo-Man

Seoul National University of Science and Technology

요약

ML(machine learning) 기술을 활용하여 실용적인 측면에서 일반 사용자들이 바라보고 사용할 수 있도록 다양한 연구 개발이 이루어지고 있다. 특히 최근 개인 사용자의 personal computer와 mobile device의 processing unit의 연산 처리 속도가 두드러지게 빨라지고 있어 ML이 더 생활에 밀접해지고 있는 추세라고 볼 수 있다. 현재 ML시장에서 다양한 솔루션 및 어플리케이션을 제공하는 툴이나 라이브러리가 대거 공개되고 있는데 그 중에서도 Google에서 개발하여 배포한 'Mediapipe'를 사용하였다. Mediapipe는 현재 'android', 'IOS', 'C++', 'Python', 'JS', 'Coral' 등의 환경에서 개발을 지원하고 있으며 더욱 다양한 환경을 지원할 예정이다.

이에 본 팀은 앞서 설명한 Mediapipe 프레임워크를 기반으로 Machine Learning을 사용한 image processing를 통해 일반 사용자들에게 편의성을 제공할 수 있는 알람 프로그램을 연구 및 개발하였다. Mediapipe에서 신체를 landmark로 검출하게 되는데 이를 scikit-learn 머신러닝 라이브러리를 사용하여 특정 자세를 학습시키고 모델화하여 알람 프로그램에 특정 기능에 조건으로 사용될 수 있게 하였다. scikit-learn은 아나콘다 등과 같은 개발환경 패키지에서 간단하게 이용 가능하는데 이 아나콘다는 데이터 분석이나 그래프 그리기 등, 파이썬에 자주 사용되는 라이브러리를 포함한 개발환경이라고 할 수 있다.

하여 본 팀은 ML기반의 영상처리 알람 프로그램을 제작하는데 있어 이러한 사항들을 파이썬 환경에서 기본적으로 포함되어 제공하는 tkinter GUI툴을 사용하고 추가적으로 인텔에서 개발한 실시간 컴퓨터 비전을 목적으로 한 프로그래밍 라이브러리 OpenCV와 여러 항목을 사용하여 환경을 구축할 수 있도록 연구, 개발하였다.

1. 제작 동기

5G나 Machine Learning, 또는 자율 주행과 같은 기술들은 중앙 처리장치에서 처리해야 하는 연산이 일반 장치에서 작동하기에는 큰 값을 가지고 있어 실생활에서는 쉬이 사용되지 않았으나 반도체 기술의 발전으로 processor의 처리 속도가 월등히 빨라지고 시장 가격도 하락하여 더 다양하고 많은 분야에서 쉽게 찾아볼 수 있게 되었다. 특히 Digital Image Processing(디지털 이미지 프로세싱)은 여러 가지 분야에서 다양하게 활용되고 있다.

이러한 Digital Image Processing(디지털 이미지 프로세싱)을 위해 인텔에서 제작한 OpenCV(Open Source Computer Vision) 라이브러리는 실시간 이미지 프로세싱에 용이하고, 윈도우, 리눅스 등 다양한 환경에서 사용 가능하여 적극적으로 사용되고 있다. 또한, TensorFlow, PyTorch 등의 Deep learning framework(딥러닝 프레임워크)도 지원하고 있어, ML(Machine Learning)을 활용하는 Digital image processing(디지털 이미지 프로세싱)에 용이하다. 프레임워크란 프로그램 개발 간에 필요한 여러 가지 라이브러리나 모듈 등 필요한 것

들을 효율적으로 접근하고 사용할 수 있도록 한 군데 모아놓은 하나의 패키지라고 할 수 있는데 이는 개발자로 하여금 중복적이고 소모적인 작업 요소를 줄여주고 기술 개발의 핵심적인 알고리즘 개발에만 효율적으로 생각할 수 있도록 도와준다.

본 연구에서 사용된 프레임워크 중에는 MediaPipe라는 것이 있는데 MediaPipe는 google사에서 제작, 배포하는 live ML 프레임워크로 Machine Learning Pipeline을 통해 multi modal과 cross platform을 구축하기 위한 프레임워크에 해당한다.[1] multi modal은 전통적인 텍스트 외에 음성, 제스처, 시선, 표정, 생체신호 등 여러 입력 방식을 융합하여 인간과 컴퓨터 사이에 자연스러운 의사소통이 가능한 사용자 친화형 기술이며, cross platform은 android, iOS, Python 등 다양한 환경에서 동작 가능한 것을 말한다. 이러한 특징들을 가지는 MediaPipe는 사용자 친화적(User-friendly) 개발환경을 가지고 있다는 장점이 있다. MediaPipe에서는 Face Detection, Face Mash, Iris, Hands 등 16가지 가량의 ML solution을 제공하고 있는데 이름에서 알 수 있듯이 얼굴 검출, 눈동자 검출, 손마디 인식 등 신체의 다양한 부위와 각각의 다른 크기의 영역에서 작동하는 솔루션들이다. 이 솔루션들은 특정

부위가 가려져 카메라가 인식할 수 없는 위치에 있어도 Deep Learning을 통해 만들어진 모델을 토대로 대략적 위치를 파악하는데, 이 대략적 위치가 거의 정확한 위치의 좌표를 보여주는데 핵심이 있다.

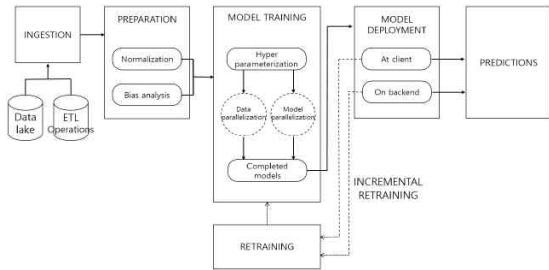


그림1. mediapipe pipeline[2]

최근 이러한 동향에 따라 실생활에 밀접하여 사용자에게 도움을 줄 수 있는 프로그램을 디자인하는 것에 초점을 두고 연구의 주제를 선정하였다. 수면 상태의 인간은 가령 무의식 상태라고 하는데 이러한 상태의 인간은 대부분의 감각이 둔해지고 시각은 거의 차단되게 된다. 때문에 알람은 스스로를 통제할 수 없는 상태에 놓인 인간에게 믿음만한 기상 솔루션을 제공해야한다. 그러나 기존의 알람 시스템은 일회성 음향 재생으로 사용자의 기상을 유도한다는 단순한 동작 방식으로 작동하는데 이러한 방식은 알람이 가져야 할 강제성과 다회성, 판단력이 부족하다고 생각하였다.

본 연구에서는 알람에 영상처리 기술에 접목을 시켜 사용자에게 체조와 같은 동작을 통해 알람을 제어함으로써 보다 강력하고 효과적인 알람 프로그램을 만드는 것을 기획을 하게 되었다. 또한 이것은 모션인식 기술에 새로운 바람을 불어넣고 나아가 알람 체계에 새로운 패러다임을 제시하여 학생, 회사원, 일반인에 관계없이 기상이 더 쉬워질 것으로 사료되며 늦잠으로 야기된 지각으로 인해 업무나 생활에 문제가 생기거나 익일 아침 세 시간에 기상해야 한다는 긴장감에 불안한 수면을 하는 일이 비약적으로 줄어들 것으로 판단된다.

2. 설계 및 구현

다른 알람과 차별화되는 부분이 바로 단순한 작동 방식이 아닌 특별한 작동 방식을 통해 알람이 울리고 정지되는 것인데 이때 가장 중요한 역할을 수행하게 되는 것이 바로 알람을 끄게 하는 조건이다. 이 조건에 특별한 방식을 적용하여 조건이 충족되게 되면 알람이 종료되는데 그러한 조건의 방식으로 영상처리를 채택했고 영상처리를 하기 위해 가장 적합한 프레임워크로 Mediapipe를 사용하였다. 또한, 그 중에서도 motion detection(동작 검출)에 가장 효과적인 holistic 모델을 핵심 solution으로 사용하였다.

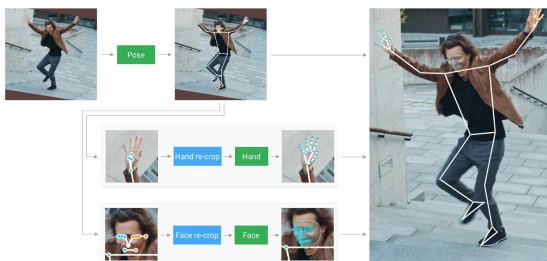


그림2. mediapipe holistic model example[3]

holistic model에서 검출한 신체부위를 landmark라고 명명하고 알람 프로그램에서 사용할 motion을 detect할 수 있도록 이 landmark의 위치에 대한 machine learning을 수행하였다. 이 때 machine learning의 프레임워크로는 scikit-learn을 사용하였다.

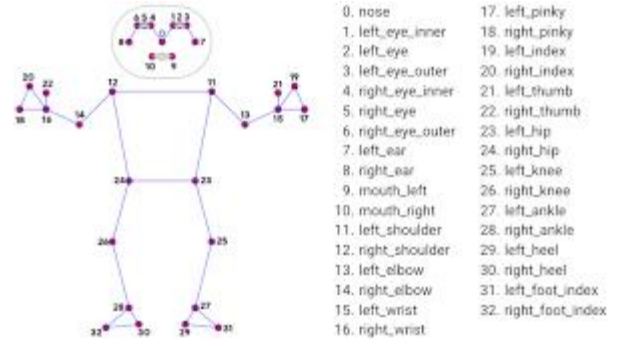


그림3. holistic model landmark[4]

scikit-learn은 Python-based machine learning 라이브러리로, 간단한 데이터 마이닝과 ML을 하는 데에 있어 쉽고 빠른 성능을 보여준다. 라이브러리 자체에 샘플 데이터 셋과 여러 가지 학습 모듈을 가지고 있어, 간편하게 원하는 모델을 활용하여 ML을 실행할 수 있다.[5] 그 중에서 linear model인 Logistic Regression(lr)과 Ridge Classifier(rc), ensemble model인 Random Forest Classifier(rf)과 Gradient Boosting Classifier(gb) 4가지를 사용하여 그 중에서 가장 높은 정확도를 보여주는 model을 채택하였다.

```
lr 0.99090909090909091
rc 0.99242424242424242
rf 0.9954545454545455
gb 0.99393939393939393
```

그림4. lr, rc, rf, gb model 각각에 대한 정확도

작품은 Python 패키지에 포함되어있는 GUI툴 tkinter를 사용하여 기본적인 알람 설정창을 만들었다. 각각 시간과 분, 초를 입력하여 시간을 받아들 수 있는 Text box를 세 개 만들고 버튼도 두 개 만들어 각각 알람 설정과 해제를 할 수 있게 하였다. 또한 직관적인 인터페이스 구축을 위해 현재 시간을 받아와 Text box와 버튼이 있는 GUI에 올려주어 시계 기능도 구현하였다. 알람이 울릴 경우 재생될 파일은 wav 파일로 미리 지정된 폴더 디렉토리에 import하였다. 그리고 모바일 기기에서 받아와 웹캠으로 인식시킨 영상을 OpenCV 라이브러리를 사용하여 사람의 동작을 detection하게 된다. 이때 미리 설정한 특정한 동작이 감지가 되면, 알람 해제의 조건(trigger)가 되어 정지할 수 있도록 설계하였다.

그러나 이 trigger는 모든 알람의 조건을 해제시키는 것은 아니다. 실제로는 알람으로 인해 재생되는 음악의 재생만을 정지 시키며 10분의 타이머를 작동시킨다. 이 타이머의 작동이 멈췄을 때 프로그램에서 받아온 영상에서 객체가 인식이 된다면 사용자가 다시 취침 상태에 빠진 것으로 판단하여 다시 알람이 울리게 된다. 이 경우 역시 앞서 취한 방식과 같은 방식을 취하게 되는데 특정한 동작이 감지되고 다시 알람이 해제되

되고 타이머 역시 다시 작동하게 된다.

이렇게 두 번째 타이머가 끝나음에도 객체가 인식이 되거나 앞선 과정에서 특정한 동작을 취하지 않게 된다면 마지막 단계로 넘어가게 되는데 이 마지막 단계는 사용자가 깊은 수면 상태에 빠졌거나 기상 의지를 상실한 것으로 판단하여 연락처에서 미리 지정해놓은 사람에게 메시지를 전달하게 된다. 이 메시지는 현재 수면에서 빠져나오지 못하는 사용자의 상태를 설명하게 되며 가능한 방법 내에서 사용자의 기상을 도울 것을 요청하는 내용을 포함하고 있다.

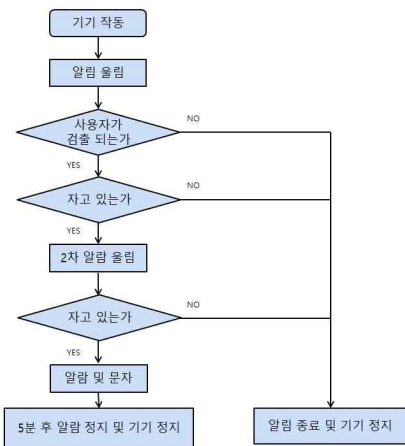
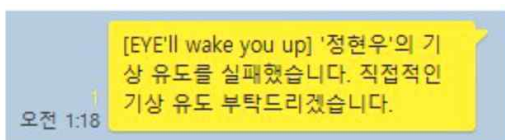


그림5. 작품 구현도 및 동작과정

이러한 방식을 통해 기존의 알람에 비해 더욱 효과적이고 유용한 프로그램을 구현하도록 노력하였다. 이러한 방식은 사용자 스스로의 의지 뿐만이 아니라, 주변인에게 메시지를 보내는 방식의 주변환경에도 강제성을 띠므로써 불규칙한 수면패턴을 가지고 아침에 힘든시간을 보내는 현대인들에게 강력한 솔루션으로 제공될 수 있는 프로그램을 만들도록 했다.

3. 구현 결과

작품의 효과를 검증하기 위해 실제로 알람 설정을 하고 취침해본 결과, 기존의 알람은 누운 자리에서 그대로 정지가 가능해 수면상태에서 각성하기 까지 시간이 오래걸렸던 반면에 위 알람 프로그램은 일어나서 동작을 취해야만 정지가 되어야 하는 강제성으로 인해 좀 더 효과적으로 기상을 할 수 있었다. 또한, 기존 알람 프로그램은 실수로 알람을 정지한 뒤에 다시 취침에 드는 실수가 생길 수 있었지만, 작품의 알람 프로그램은 계속해서 기상상태를 체크하고서도 모션을 취하지 않으면 메시지를 통해 주변에 알릴 수 있기 때문에 다시 잠드는 실수를 면할 수 있고, 혹여나 다시 잠에 들더라도 주변에서 이를 확인하고 연락을 취할 수 있는 등 효과적인 결과를 보여주었다.



<최종적으로 보내지게 되는 메시지>

그림6. 최종적으로 보내지는 메시지

작품을 제작하는 과정에서 영상처리를 반복작업을 해야하는 것에서 최대한 프레임저하가 발생하지 않고, 빠른 연산이 이루어지기위해 비교적 high-level의 프레임워크들을 최대한 이용하였다. 또한 ML도 비교적 큰 동작구분을 이용해서 조금 다양한 환경과 각도에서도 정확한 결과값을 도출 할 수 있도록 설계를 하였다. 이러한 것들이 테스트를 하는 과정에서는 크게 문제가 발생되지는 않았지만, 실제 제품으로써 상용화가 되기 위해서는 조금 더 정밀도를 높이고 최적화가 되어야 할 것으로 보인다.

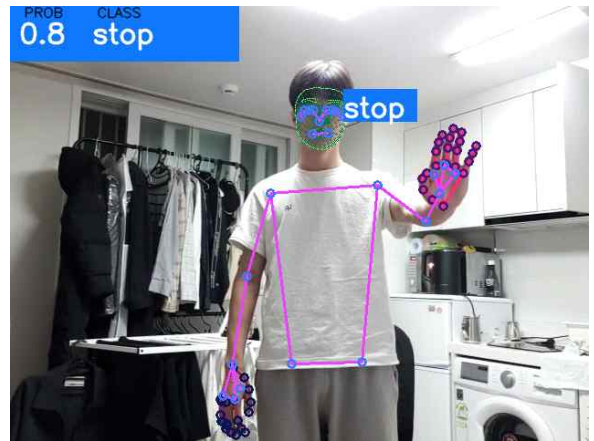


그림7. 작품의 영상처리 장면



그림8. 작품의 알람 프로그램 GUI

4. 작품의 기대효과

가장 처음 연구의 방향성을 안드로이드 환경의 모바일 어플 개발로 설정하였으나 연구에 사용된 라이브러리나 툴, 모듈 등이 모두 여러 환경에서 사용 가능한 cross platform을 제공하고 있기 때문에 프로그램을 windows PC 환경에서 선제적으로 구축했다. 가장 핵심적인 Mediapipe가 안드로이드 환경을 지원하기 때문에 추후 본 연구를 develop함에 있어서 scikit-learn을 통해 학습된 데이터를 사용하여 Java 또는 Kotlin을 사용한 안드로이드 스튜디오에서 코드를 재작성하여 간단하게 구현할 수 있을 것으로 전망한다. 현재 구현되는 방식은 windows체제의 PC와 영상을 받아오는 웹캠(mobile) 두 개의 장치를 필요로 하지만, 이를 통해 mobile기기 하나만으로 동작할 수 있는 환경을 갖출 수 있게 될 것이다.

기존에도 단순히 정지버튼만을 통해 제어하는 알람 프로그램에 문제점을 해결하기 위해 수학문제를 풀게하거나, 휴대폰을 흔드는 동작등

통해 알람을 제어하는 형식의 프로그램도 존재했었다. 하지만 실시간 영상처리를 통해 제어하는 형식의 프로그램은 존재하지 않았고, 이러한 형식은 기존의 프로그램들 보다 훨씬 더 강력한 효과를 가지고 있다고 보여진다. 현재는 특정 동작을 취한 것으로 인식을 하는 것에서 그쳤지만, 위의 형식을 이용하여 연결된 동작을 통해 실제로 체조와 같은 것을 트리거로 사용한다면, 훨씬 더 효과적이고 능동적인 기상효과를 불러올 수 있을 것으로 보인다.

이번 작품을 만드는 데에 있어 알람간격, 알람음악, trigger motion 등의 변수는 임의로 지정하여 사용하였는데, 이것을 실제 알람 프로그램으로 상용화를 한다면, 사용자 지정이 가능하도록 interactive 한 코드로 재작성을 하여 개발할 수 있을 것이다. 또한, 계속해서 반복적인 영상처리를 하는 과정에서 불가피하게 프레임저하가 조금씩 일어나 기도 했는데, 이러한 점들을 개선한다면 더욱 완성도 높은 결과물이 나올 것으로 보인다.

본 작품은 실생활에서 빼놓을 수 없는 기능 중에 하나인 알람기능에서 출발하여 현재 상용화되어 있지 않은 영상처리를 통한 새로운 방식의 알람 프로그램의 형태이다. 이는 서론에서 언급한 것과 같이 더욱더 mobile기기의 성능이 높아지면서 충분히 적용이 가능한 강력한 알람기능의 방식으로 개발가능한 부분을 다루어보았다고 생각한다.

5.

[1] Mediapipe webpage ('google') :

<https://google.github.io/mediapipe/>

[2]<http://blog.skby.net/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-%ED%8C%8C%EC%9D%B4%ED%94%84%EB%9D%BC%EC%9D%B8-machine-learning-pipeline/>

[3]<https://google.github.io/mediapipe/solutions/holistic.html>

[4]<https://google.github.io/mediapipe/solutions/pose.html>

[5] scikit-learn, Machine Learning in Python :

<https://scikit-learn.org/stable/>