

딥 러닝 기반의 API 와 멀티미디어 요소를 활용한 시니어 라이프 데이터 수집 및 상태 분석

김선대, 박은수, 정종범, 구자성, 류은석
가천대학교

{ele7004, dmseh804, uof4949}@gc.gachon.ac.kr, 32122610@dankook.ac.kr,
esryu@gachon.ac.kr

Senior Life Logging and Analysis by Using Deep Learning and Captured Multimedia Data

Seon Dae Kim, Eun Soo Park, Jaseong Koo, Jong Beom Jeong, Eun-Seok Ryu
Gachon University

요 약

본 논문에서는 시니어를 위한 라이프 데이터 수집 및 행동분석 프레임 워크를 설명하고, 이의 부분적 구현을 자세히 설명한다. 본 연구는 시니어를 위한 라이프 데이터를 바탕으로 보호자가 없는 시니어를 보살핌과 동시에, 보호자가 미처 인지하지 못하는 시니어의 비정상적인 상태를 분석하여 판단하는 시스템을 연구한다. 먼저, 시니어가 시간을 많이 소요하는 TV 앞 상황을 가정하고, 방영되는 TV 콘텐츠와 TV 카메라를 이용한 시니어의 영상/음성 정보로 이상상태와 감정상태, TV 콘텐츠에 대한 반응과 반응속도를 체크한다. 구체적으로는 딥 러닝 기반의 API 와 멀티미디어 데이터 분석에서 사용되는 오픈 패키지를 바탕으로, 영상/음성의 키 프레임을 추출하여 감정 및 분위기를 분석하고 시니어의 얼굴 표정 인식, 행동 인식, 음성 인식을 수행한다.

1. 서론

UN(United Nations)는 국내에 65 세 이상 비율이 14% 이상이면 고령사회, 20% 이상이면 초고령사회로 분류한다. 최근 우리나라는 2017 년 8 월말 주민등록 인구수가 약 5,175 만명으로 고령사회에 진입하였다. 또한 KOSIS 국가통계포털에 따르면, 2014 년 기준 노인 우울증은 65 세 이상의 인구에서 3 명중 1 명꼴로 우울증을 앓고 있다[1]. 우울증 군에 속한 시니어는 일상생활에서나 인지기능에서 비우울증 군과 유의미한 차이를 보이고, 치매의 위험이 높기 때문에[2], 우울증과 치매로 인한 국가와 가정의 의료비 부담이 더욱 늘어날 수밖에 없다.

본 논문은 그림 1 과 같이 시니어를 보살핌 수 있는 라이프 로깅을 통해 시니어의 상태를 자동적으로 실시간 관찰하여, 그 결과를 분석하여 판단하고 보호자나 주치의에게 알리는 시스템을 연구한다. 관련 연구에서는 기존에 진행되었던 라이프 로깅에 대한 연구를 소개한다. 본론에서는 본 논문에서 라이프 로깅 및 분석/판단을 연구한 내용에 대해 서술한다.

본 논문에서 연구하는 비디오 및 오디오 콘텐츠는 TV 에서 추출되는 영상이나 음성 데이터를 말한다. 이 두 가지 데이터는 시니어의 상태를 주기적으로 분석하고 판단하는 데 사용된다. 예를 들어 공포영화에서 무서운 장면이 방영될 시에 사람이 행복해지는 모습을 보이거나 웃고 있다면 정상적인 사람의 반응이 아니라고 여겨진다. 즉, 콘텐츠에 대응하는 시니어의 반응이 콘텐츠에서 기대되는 반응과 달라 시니어의 정신상태가 정상 또는 비정상인지 예측 및 판단을 할 수 있는 중요한 데이터가 될 수 있다. 추가적으로, 일반적인 TV 에서 시청하는 콘텐츠의 제목과 시간은 방송 콘텐츠에서 추출이 가능하며, IPTV 의 경우 콘텐츠의 장르, 평점 등의 추가적인 정보 추출이 가능하다. 이 정보들은 물론 콘텐츠의 전체적인 분위기 파악에 유용하지만, 공포영화의 중요 프레임마다 부정적인 감정으로 ‘공포’ 인 장면이 있을 수 있고, 반대로 긍정적인 감정으로 ‘행복’ 의 감정인 장면이 나올 수 있다. 이렇듯 장르 기반으로만 콘텐츠를 규정하면 시니어가 행복한 장면에서도 웃고 있지만 비정적인 결과로 도출될 수 있다.



그림 1. 본 연구가 제안하는 시니어 라이프 데이터 수집 개념도

이를 보완하기 위해 콘텐츠의 영상을 중요 프레임 별로

나누어 딥 러닝 기반의 오브젝트 탐지, 행동 인식, 이미지 필터링, 얼굴 및 표정 인식, 이미지 캡처 등 여러 센싱 기술들을 이용하여 각 프레임의 특징을 추출하고 분위기를 판단한다. 또한, 시니어의 라이프 데이터를 더욱 구체적으로 모으기 위해 수행한 보조연구로서 (1) 웨어러블 기기를 통한 시니어의 심박수, 급격한 행동변화 등을 측정한다. (2) 치매 및 우울증 상태감사지를 통한 신체/정신상태 데이터를 얻고, 이를 활용한다.

2. 관련 연구

2.1. 깊이 지도 기반의 사람 행동 추적 및 라이프 로깅

시니어의 행동인식을 하기 위한 방법은 시니어의 거주지에서 간단한 연산능력이 있는 에지카메라(Edge Camera)로 캡처하거나, 카메라로 촬영한 뒤 서버에서 데이터를 처리하는 방법이 있다. 이에 관련된 연구[3]는 주택이나 건물과 같은 생활 환경에서 인간의 활동을 기록하는 스마트 라이프 로깅을 한다. 이는 건강 관리에 사용될 수 있는 거주자의 일상 생활 활동을 지속적으로 모니터링하고 기록한다. 이에 라이프 로깅 방법을 보다 자세하게 설명한 [4]에서는 2차원 깊이 지도를 이용하여 인간 활동 추적 및 인식 시스템을 구현한다.

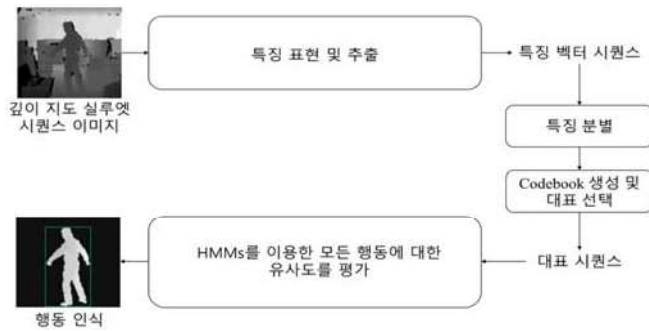


그림 2. 깊이 지도를 통한 특징 추출 및 행동 인식 흐름도[4]

그림 2 와 같이 깊이 카메라를 통해 만들어진 깊이 지도는 움직임 시퀀스의 일시적 연속성 제약을 고려하여 사람의 실루엣을 추적한다. 특징 추출에서는 차원 감소(연산량 감소 및 정확도 향상)를 위한 Principle Component Analysis(PCA), 특징을 더욱 잘 구별하기 위한 Linear Discriminant Analysis(LDA)를 이용하였고, 마지막으로 Hidden Markov Models(HMMs)를 이용해 사람의 행동을 인식한다[4].

2.2. 다양한 센서들을 활용한 라이프 로깅

최근에 수행된 연구인 [5]에서는 정신 질환을 관찰 및 분석하는 감지 기술에 대해 종합적으로 설명한다. 보통 사람들이 많이 이용하는 전자기기에 부착된 센서들(가속 센서, 자이로 센서, 심박 센서), 어플리케이션(통화, 문자메세지, 소셜네트워크 어플리케이션)등을 기반으로 사용자의 정신 질환을 분석한다. 센서의 신호는 크게 3 가지로 행동 신호, 생리적 신호, 사회적 신호를 이용하며, 이는 정신 질환인 우울증, 조현병, 조울증, 자살 충동, 불안장애, PTSD(외상성 스트레스 장애) 등의 정신질환과의 관계를 해석한다. 동시에 사생활 침해, 제한된 실험환경, 데이터 수집 기술, 센서 기술의 무결성 등과 같은 라이프 로깅 연구에 대한 한계를 설명한다.

그리고 [6]에서는 앞서 [5]에서 설명한 바와 같이, 다양한 센서들을 통해 파킨슨병 환자와 알츠하이머병 환자들의 행동을 분석하여 보살피는 연구를 수행한다. 아래의 그림 3 과 같이 의료데이터, 센서기기를 통한 건강상태 데이터, 360/깊이

카메라 등을 활용한 비정상 행동을 감지하는 복합적인 인지 데이터를 토대로 환자의 데이터를 수집 및 분석한다. 시각, 움직임, 깊이 센서들을 토대로 환자가 생활하는 거주지, 노인 센터, 재활 센터 등에서 연구를 수행한다. 이는 [5]에서 설명한 라이프 로깅 연구의 한 방법으로, 환자의 생활환경 내의 행동 및 움직임을 분석하고 판단한다. 환자의 생활권 내의 궤도를 측정하여 환자의 상태를 체계적으로 분석하는 것을 가능케 한다. 또한 55 세 ~ 94 세 사이의 18 명의 정신 질환 환자들을 10 주 동안 실측하여 테스트를 수행한다[6].

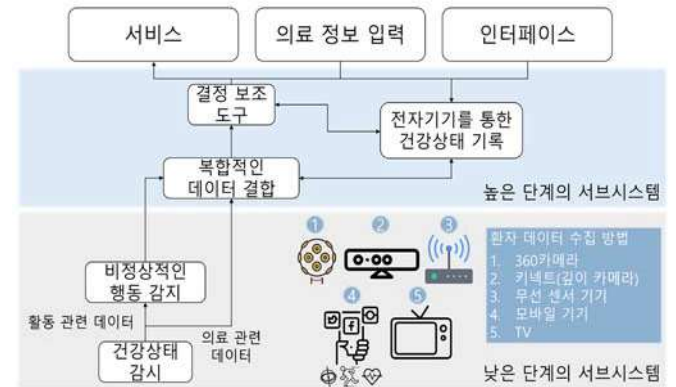


그림 3. 환자의 행동 분석을 위한 행동 감지 시스템 구조도[6]

3. 멀티미디어를 활용한 시니어 라이프 로깅

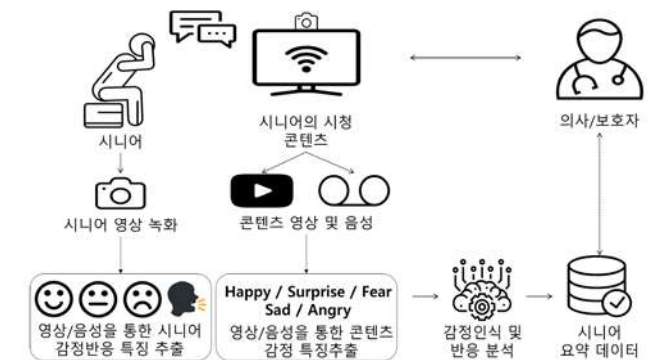


그림 4. 본 연구의 시니어 라이프 로깅 연구개발 개념도

본 연구는 그림 4 와 같이 TV 콘텐츠, 시니어 캡처 영상 및 음성 데이터를 활용한 시니어 라이프 분석을 수행한다. 시니어가 거주지 내에서 시간을 많이 보내는 장소인 TV 앞(거실 또는 방)을 고려하며, TV 에서 방송되는 콘텐츠와 에지카메라를 통해 시니어를 촬영한 모습을 기록한다. 이는 딥 러닝 기반 API 를 통해 감정을 분석하고 판단한다. 또한 감정 특징 분석을 통해 콘텐츠에 어울리지 않는 시니어의 이상반응을 측정한다. 이때 콘텐츠와 시니어를 캡처한 영상을 모두 판단하기에는 에지카메라의 연산능력에 있어 오버헤드를 가져오므로, 본 연구는 시니어를 캡처한 영상에서 일정수준 이상 표정, 행동, 음성이 변화하는 것을 기준으로 중요 프레임을 선정한다[12].

3.1. 시청하는 비디오 및 오디오 콘텐츠 분석

아래의 그림 5 와 같이 이미지 캡처와 오브젝트 탐지, 표정 인식은 콘텐츠의 감정 추출과 시니어의 행동인식 및 감정인식에 매우 중요하게 쓰일 수 있다. 콘텐츠에서의 주인공의 행동 또는 오브젝트의 흐름에 따라 대부분의

분위기가 결정되기 때문에, 그와 직결되는 중요 프레임에 대한 감정 추출에 유용하게 쓰이기 때문이다.

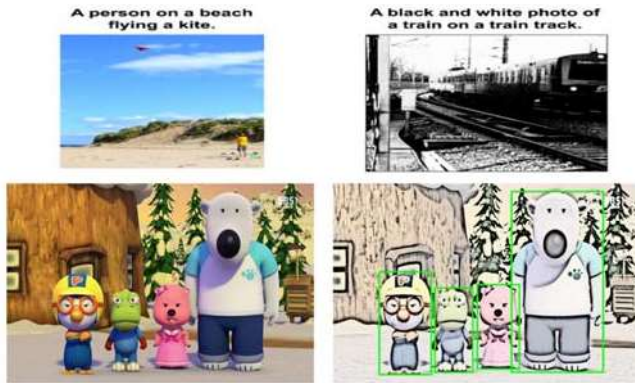


그림 5. 이미지 캡셔닝(위의 두 사진)과 오브젝트 탐지의 예(아래의 두 사진)

음성 데이터는 음성을 텍스트로 변환하여(Speech to Text)를 이용하여 텍스트로 바꾼 다음, 각 감정에 해당하는 단어를 학습한 모델로 판단하는 방법과 각 감정에 해당하는 음성 데이터를 레이블로 지정하여 특징을 추출/학습하는 방법이 있다. 두 가지 병합하여 모두 사용하는 방법은 실측 감정과 더욱 유사한 결과가 나올 수 있다. 하지만 음성의 텍스트변환 성능이 좋지 못하면, 감정 분석결과 자체가 부정확해질 수 있다. 그리고 사용 언어가 다를 경우 텍스트 변환이 더욱 어렵게 된다. 이를 위해 음성의 텍스트변환 성능이 뛰어난 API 인 CSR(Clova Speech Recognition) [7]와 음성 데이터 감정분석 오픈소스[11]를 함께 이용한다.

3.2. 시니어 캡처 데이터

본 연구의 주요 목적인 시니어 라이프 로깅을 통한 시니어의 상태 분석은 예지 카메라를 통해 시니어를 캡처하고 상태를 요약적으로 기록하는 것이 매우 중요하다. 이에 시니어를 캡처한 데이터는 얼굴, 행동, 음성데이터로 총 3 가지의 종합적인 인식 및 분류를 통한 시니어의 상태를 분석 및 판단한다. 캡처한 영상은 중요 프레임 별로 나누어 시니어의 상태를 분석한다[12]. 먼저, 시니어의 얼굴 이미지 캡처는 각각의 영상 프레임에서 얼굴을 따로 잘라내어 이미지로 저장한다. 이러한 얼굴과 표정인식은 비교적 높은 인식률을 보이는 [7]-[9]을 사용한다. 이는 얼굴을 인식하고 얼굴 이미지를 추출하여 표정 분석을 통한 감정 추출을 수행한다. 행동인식은 사람의 신체의 상태를 인식한다. 서있거나, 앉아있거나, 누워있는 등의 고정된 자세와 움직이는 모습을 추출하는 데 목적이 있다. 이에 본 연구에서는 행동인식에서 높은 인식률을 보이고 있는 [10]를 이용한다. 각 영상 프레임에서 신체를 인식하고 관절의 포인트와 뼈대를 이용한 스켈레톤 형식으로, 몸과 사지의 위치, 각도 등을 특징으로 추출할 수 있다. 이는 [3]-[6]와 같이 거주지에서 활동하는 사람의 행동을 실시간 모니터링 하고 행동 패턴을 인식할 수 있다. 마지막으로, 음성인식은 콘텐츠의 음성을 추출하여 특징을 통해 분석 및 판단하듯이, 시니어의 캡처 데이터에서 음성을 추출하여 같은 방법을 사용한다.

추가적으로, 본 연구에서는 웨어러블 디바이스를 사용하여 행동 인식 및 신체상태 정보수집을 수행한다. 웨어러블 디바이스는 사람의 심박수, 행동 변화, 걸음걸이 수 등의 헬스케어 목적으로 많이 이용되어 시니어의 건강상태를 측정하는 데 유용하다. 특히 웨어러블 디바이스를 착용한

시니어는 수면 상태를 체크할 수 있다. 수면의 질이 좋지 않을 시에 치매에 걸리게 될 확률이 더욱 높아지기 때문에[13], 그 전에 문제를 감지하여 치료할 수 있다. 더불어, 시니어의 상태를 객관적으로 평가할 수 있는 지표인 노인 치매 및 우울증에 대한 검사 신뢰도가 높은 상태감사지를 활용한다. 문제의 답변이 오래 걸릴수록 시니어의 인지기능에 문제가 있을 수 있어, 그 반응시간을 측정한다[14].

4. 실험 및 구현

멀티미디어 콘텐츠와 시니어 영상 및 음성 데이터는 오픈 패키지나 오픈 API 를 이용하여 분석한다. 비음문제와 정확도를 감안하여 그림 6 과 같이 본 연구는 네이버사의 Clova CFR (Clova Face Recognition) API[7]를 이용하여 얼굴인식을 구현한다. 그리고 얼굴을 캡처 화면에서 분별하고 얼굴 이미지를 따로 저장하기 위해 OpenCV [8],[9]를 이용한다.

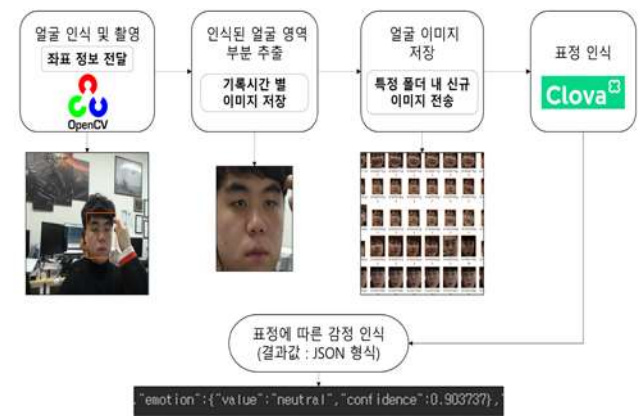


그림 6. OpenCV 와 CFR 을 활용한 얼굴 및 표정 인식 흐름도

아래의 그림 7 과 같이, 행동 인식은 촬영되는 영상의 사람에게 신체 포인트를 적용하여 스켈레톤을 구현하는 OpenPose 오픈소스[10]를 이용해 구현한다. 행동, 얼굴의 모양, 표정도 추출할 수 있으며 손으로 표현하는 다양한 모션도 추출이 가능하다. 또한 웨어러블 디바이스인 삼성전자의 갤럭시 기어 S3 는 블루투스 와 와이파이 등으로 연결할 수 있는 SAP(Samsung Accessory protocol)을 통해 스마트폰과의 정보전달을 구현할 수 있다. 갤럭시 기어는 구동 운영체제인 타이젠(TIZEN)의 SDK 와 Android Studio 3.0 기반의 어플리케이션을 통해 구현한다.



그림 7. 사람의 행동을 측정하는 OpenPose(위)와 웨어러블 기기를 통한 센서 정보전달 테스트(아래)

마지막으로, 음성인식은 그림 8 과 같이 안드로이드 기반의 CSR [7] API 를 이용하여 시니어의 음성을 높은 정확도로 기록한다. 또한 푸리에 변환(Fourier Transform)과 SVM(Support Vector Machine)을 사용하여 소리를 분별/분석하고 판단하는 리눅스 기반의 음성인식 기술 [11]을 사용한다. 음성 텍스트 변환, 푸리에 주파수 변환을 바탕으로 텍스트 및 음성 신호의 특징을 추출하여 SVM 분류기로 감정을 분석 및 판단한다. 음성에 대한 감정은 그림 8 하단의 사진과 같이 잘 인식할 때도 있는 반면 (Angry>Angry), 긍정/부정의 감정을 같은 부류의 다른 결과로 나타낼 때도 있다(Fear>Angry). 이는 3.1.에 제시한 음성의 텍스트 변환 분류 성능이 좋지 못한 것이 가장 큰 이유이다.

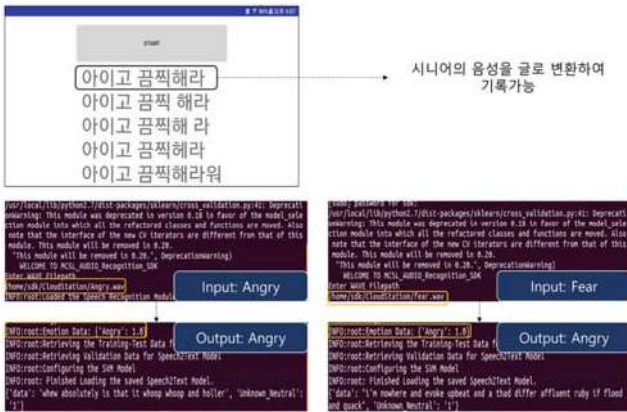


그림 8. 음성 텍스트 변환(위)과 음성 데이터를 통한 감정 추출 (아래)

5. 결론

본 논문에서는 시니어를 위한 딥 러닝 기반의 API 와 멀티미디어 요소를 활용한 라이프 로깅과 상태 분석을 수행한 연구 내용을 서술하였다. 본 논문에서는 크게 행동인식, 표정인식, 음성인식 총 세가지로 나누어 시니어 라이프 로깅 시스템을 서술하였다. 본 연구는 시니어를 위한 라이프 로깅을 위해 딥 러닝 기반 오픈 API 및 오픈소스를 활용하여 표정에서 감정을 추출하고, 행동인식을 위한 스켈레톤 연구, 웨어러블 기기의 센서 연구, 음성을 통한 감정인식 연구를 수행한다. 본 논문의 한계는 시니어에 대한 데이터가 매우 부족하다는 점과 영상이나 음성에 대한 감정 추출에 대한 연구가 추가적으로 수행되어야 한다는 점이다. 표정인식이나 음성에 대한 감정추출 연구는 많이 진행되어 왔지만, 영상 콘텐츠에 대한 감정 추출은 조금 더 심도 있는 연구가 필요하다. 특히 그림 7 과 같이 음성 감정추출의 경우 따로 텍스트와 음성 감정으로 분류하여 기록하는 이유는 한국어의 경우 텍스트로 기록하는 연구는 많지만, 감정을 나타낸 음성 데이터셋과 감정을 추출하기 위한 연구가 부족하기 때문이다. 추후 연구로써 (1) 시니어 라이프 데이터 확충 및 TV 콘텐츠 추출과, (2) 1 차적인 영상, 음성, 웨어러블 센싱 정보에 대한 각각의 딥 러닝 판단을 활용하여 2 차적으로 종합/복합적으로 최종 상황판단을 내리는 복합지능 프레임워크로의 확대 연구가 있다.

Acknowledgement

본 연구는 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음. [GRRC-가천 2017(B01), 시니어 라이프 로그 기반 행동 분석]

참고문헌

- [1] 통계청 국가통계포털, “KOSIS 100 대 지표”, 2014, (http://kosis.kr/conts/nsportalStats/nsportalStats_0102Body.jsp?menuId=11&NUM=1100)
- [2] 손석한, et al. “알츠하이머형 치매 환자에서 우울증의 유무에 따른 인지기능과 일상생활 수행능력의 차이.” J Korean Neuropsychiatr Assoc 38.2. 1999.
- [3] Jalal, Ahmad, et al. “Development of a life logging system via depth imaging-based human activity recognition for smart homes.” Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea. Vol. 19. 2012.
- [4] Jalal, Ahmad, et al. “Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home.” IEEE Transactions on Consumer Electronics 58.3. 2012.
- [5] Abdullah, Saeed, et al. "Sensing Technologies for Monitoring Serious Mental Illnesses." IEEE MultiMedia 25.1: 61-75. 2018.
- [6] Alvarez, Federico, et al. "Behavior Analysis through Multimodal Sensing for Care of Parkinson's and Alzheimer's Patients." IEEE MultiMedia 25.1: 14-25. 2018.
- [7] Clova. “Naver Developers Clova A.I. APIs”, 2018, (<https://developers.naver.com/docs/clova/api/>)
- [8] Bradski et al. “OpenCV. Dr. Dobb’s journal of software tools 3”, 2000.
- [9] OpenCV. “Open Source Computer Vision Library”, 2018, (<https://opencv.org/>)
- [10] Cao, Zhe, et al. “Realtime multi-person 2d pose estimation using part affinity fields.” CVPR. Vol. 1. No. 2. 2017.
- [11] Vyas Kovakkat, “DeepSentiment: Speech Emotion Recognition using FFT and SVM.”, 2017, (<https://github.com/vyassu/DeepSentiment>)
- [12] S.M. Guo, et al. “A Key Frame Selection-Based Facial Expression Recognition System” Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC’06), IEEE, 2006.
- [13] Lucey, Brendan P., et al. “Effect of sleep on overnight cerebrospinal fluid amyloid β kinetics.” Annals of neurology 83. 1: 197-204, 2018.
- [14] 고영규. “치매가 동작수행을 위한 정보처리 과정 중 반응선택 작용에 미치는 영향” Korean Journal of Sport Psychology 2002, Vol. 13. No. 1, 153~161, 2002.