

컨볼루션 신경망 기반 표정인식 스마트 미러

최성환 · 유운섭*

한경대학교

Smart Mirror for Facial Expression Recognition Based on Convolution Neural Network

Sung Hwan Choi · Yun Seop Yu*

Hankyong National University

E-mail : kkorack2@hknu.ac.kr / ysyu@hknu.ac.kr

요 약

본 논문은 여러 인공지능 기술 중 이미지 분류를 통한 사람의 얼굴 표정을 인식하는 프로그램을 통해 사람의 표정을 인식하여 거울에 나타내는 스마트미러 기술을 소개한다. 여러 사람의 5가지 표정 이미지를 통하여 인공지능으로 학습하였고, 사람이 거울을 볼 때 거울이 그 표정을 인식하여 인식한 결과를 거울에 나타내는 방식이다. 여러 사람의 얼굴을 표정별로 구분되어있는 dataset을 kaggle에서 제공하는 fer2013을 이용하여 사용하였고, 이미지 데이터 분류를 위해 네트워크 구조는 컨볼루션 신경망 구조를 이용하여 학습하였다. 최종적으로 학습된 모델을 임베디드 보드인 라즈베리파이4를 통해서 얼굴을 인식하여 거울을 통해 디스플레이에 나타내는 구조이다.

ABSTRACT

This paper introduces a smart mirror technology that recognizes a person's facial expressions through image classification among several artificial intelligence technologies and presents them in a mirror. 5 types of facial expression images are trained through artificial intelligence. When someone looks at the smart mirror, the mirror recognizes my expression and shows the recognized result in the mirror. The dataset fer2013 provided by kaggle used the faces of several people to be separated by facial expressions. For image classification, the network structure is trained using convolution neural network (CNN). The face is recognized and presented on the screen in the smart mirror with the embedded board such as Raspberry Pi4.

키워드

Smart Mirror, Convolution neural network, Edge Computer, Facial expression recognition

1. 서 론

최근 고령화 사회와 비혼 추세로 인해 1인가구가 4년 전에 비해 4%가량 증가하였고, 더 증가할 것으로 예측된다 [1]. 1인가구의 문제점은 집에 혼자 있어서 다양한 문제가 발생했을 때 대처하기 어렵고, 다른 사람과의 의사소통이 없어 외로움이 느껴진다. 언어 처리 및 이해, 컴퓨터 비전, 기계

추론, 딥러닝과 같은 여러 기능 중에 딥러닝을 통해 개발된 기기를 이용해서 1인가구의 문제점을 해결하는 노력이 진행되고 있다[2].

본 논문은 1인가구의 문제점을 해결하기 위해서 사용자의 얼굴 표정을 딥러닝으로 인식해서 사용자의 감정에 변화를 줄 수도 있게 하는 스마트미러에 대해서 소개한다.

* corresponding author

II. 스마트 미러 제안

2.1 전체 구성도

제안한 스마트미러의 구성도는 그림 1과 같다. 이 스마트 미러는 하프미러필름을 붙힌 아크릴판 뒤에 디스플레이와 임베디드 보드인 라즈베리파이 4를 배치하였고, 중간에 카메라를 설치하여 거울 앞 상황을 카메라를 통해 임베디드 보드인 라즈베리파이[3]로 전송하고 이 보드에서 표정인식과 음성출력을 제어한다.

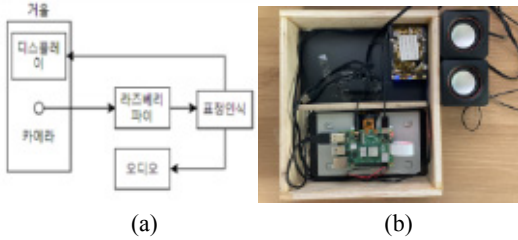


그림 1. (a) 스마트미러 구성도와 (b) 실물 사진

2.2 동작 알고리즘

임베디드 보드에 연결된 카메라로부터 촬영한 거울 앞 사람의 얼굴을 인식하게 되면 프로그램이 시작된다. 얼굴 인식 후에 카메라가 가져온 영상을 가지고 표정인식을 하게 되는데 표정을 인식 한 후 list에 그 결과 값들을 저장한다. 저장한 후 저장된 값이 일정 값 이상이 되면 저장된 값 중 가장 많은 값으로 선택하여 디스플레이한다.

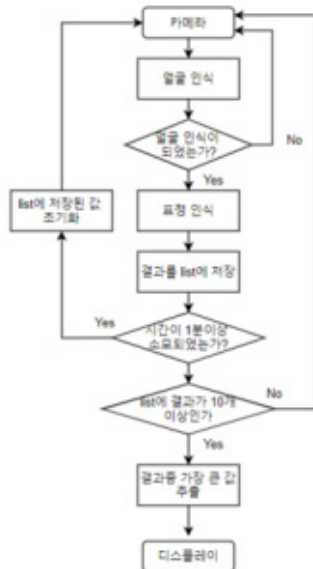


그림 2. 동작 알고리즘

2.3 얼굴인식

얼굴인식은 이 프로그램을 작동시키는 시작점이다. 카메라가 얼굴에 인식이 된 순간부터 프로그램이 시작되기 때문에 만약 거울 앞에 얼굴인식이 되지 않는다면 작동되지 않는다. 얼굴인식은 opencv에서 제공하는 Haar Cascades[4]를 사용하였다.

2.4 표정인식

2.4.1 Dataset

Dataset은 kaggle에서 제공하는 fer2013을 사용하였다[5]. fer2013의 데이터를 보면 총 7가지 특징으로 분류 되어있는데 그림 3과 같이 그 중 5가지의 특징으로 분류하였다. 이미지의 크기는 48×48이고, 데이터의 총 개수는 30217개이다. 이 데이터 중에 80%인 24173개는 훈련데이터이고 20%인 6044개는 테스트데이터로 사용하였다.

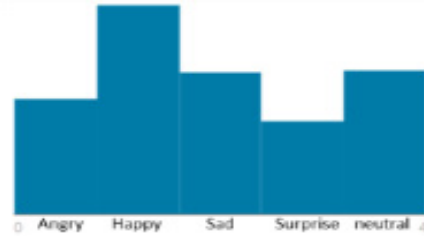


그림 3. 5가지 표정 이미지 dataset 구조

2.4.2 Network

Network 구조는 그림 4과 같이 layer_1과 layer_2의 구조가 나열된 방식이다. input 데이터가 layer_1의 구조를 총 9번 통과한 후 layer_2를 한번 통과한 구조이다. 모델을 학습시킬 때 learning rate = 0.001, batch size = 32, epoch = 1000번으로 하였다. 총 5가지 표정으로 분류하였고, 정확도는 74%정도로 나타났다.



그림 4. Network 구조

III. 결과 및 성능 비교

3.1 Network 비교

그림 5는 성능 비교를 위해서 사용된 3가지 네트워크 모델을 나타낸다. 모델 1은 layer_1이 5번 사용되고, 모델2는 layer_1이 7번 사용되고, 모델3은 layer_1이 9번 사용된 연결된 구조이다.

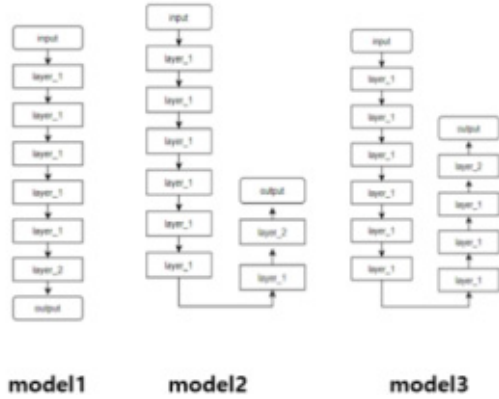


그림 5. 비교 Network

각 모델의 성능은 표1에 보여진다. Test 데이터의 정확도는 거의 비슷한 결과를 보였다. Train 데이터는 모델마다 차이를 보였는데 모델1이 가장 높은 정확도를 보였고, 모델3이 가장 낮은 정확도의 결과가 나타났다. 하지만 각각의 모델을 가지고 거울 앞에서 표정을 지었을 때는 모델3이 가장 좋은 성능을 나타냈는데, 이 결과로 보아 모델1,2가 모델3보다 Train정확도가 높은 이유는 overfitting이 발생한 것으로 보인다. 따라서 overfitting이 발생하면 실제 성능이 낮아지는 것을 알 수 있다.

표 1. Network 성능 비교

	Test	Train	실제 정확도
model1	74.189	77.7	Sad 표정이 잘 안 나타남
model2	73.594	76.2	안경을 썼을 때 surprise,sad 표정이 잘 안 나타남
model3	74.5698	74.7	안경을 썼을 때 sad 표정이 잘 안 나타남

3.2 속도 비교

표정인식을 하는데 결과가 나타나는 시간도 중요한 부분 중 하나였다. 라즈베리파이4의 성능이 일반 컴퓨터보다 떨어졌고, 그로인해 복잡한 모델을 만들수록 결과가 나타나는데 오랜 시간이 걸렸다. 하지만 그림 6과 같이 3개의 모델은 layer와 filter의 수가 많지 않아서 인식시간이 크게 차이

나지 않았다. 세 가지 모델을 비교 분석한 결과, 정확도에서 가장 좋은 결과는 모델3에서 보여진다.

	인식 시간
model1	4.5~4.7
model2	4.5~4.8
model3	4.7~5

그림 6. 인식 시간

IV. 출력 결과

표정인식이 끝난 결과를 디스플레이에 출력하는 부분은 gui를 이용하였다. 먼저 표정이 인식되는 동안에는 그림 7과 같은 문구가 나타난다.

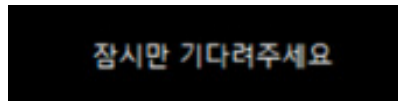


그림 7. 표정 인식하는 동안 출력

표정인식이 끝난 후 list에 가장 많은 값을 찾아 출력하는데 각각의 표정마다 다른 결과를 그림 8과 같이 나타낸다.

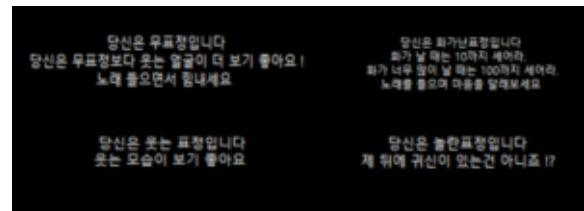


그림 8. 각 표정에 대한 결과

표정인식은 angry, surprise, sad, happy, neutral 총 5가지의 결과로 분류하였다. 각각의 표정마다 결과를 다르게 하였다. 이 중 angry, sad, neutral은 마음을 편안하게 해주는 음악을 넣어 나를 위로해 주고 공감해줄 수 있는 기분이 들게 하였고, 각각의 표정에 맞게 위로 해주는 말이나 교훈 또는 공감해주는 말을 해주면서 1인가구의 문제점을 해결하고자 하였다. happy와 surprise는 음악소리가 나오지 않지만 유머러스한 말이나 칭찬하는 말을 넣어 다른 표정보다 더 많은 결과가 나타나도록 하였다.

V. 결 론

본 논문에서는 CNN 모델로 kaggle에서 제공하는 데이터인 fer2013을 이용하여 사용하였다. 30000개의 훈련데이터와 테스트데이터로 3가지 모

델로 비교해보면서 분석을 했다. 초기에는 fer2013 데이터 뿐만 아니라 내 얼굴을 사진으로 찍어 테스트데이터에 추가한 후 결과를 비교해 보았는데 60%미만의 결과가 나타났다. 사진을 찍을 때 배경 부분을 포함한 상태로 찍었을 뿐만 아니라 앞머리가 내 이마를 전체적으로 가리고 있어서 눈썹의 특징을 지워버린 것이 문제였다. 표정인식을 할 때 사진을 데이터로 사용하려면 먼저 하얀 배경위에서 사진을 찍고, 눈썹 눈 코 입 등 얼굴 표정에 특징들이 잘 나타나도록 해야 한다.

3가지의 모델을 비교할 때 실제 정확도에서 안경을 낀 때에 결과가 나타나지 않는 경우가 많이 보였다. 학습시킨 데이터셋이 대부분 안경을 끼지 않고 있고, 안경을 끼게 되면 각 표정의 특징들 중 가리는 부분이 생기게 되고, 그 뿐만 아니라 안경이라는 새로운 특징이 추가되기 때문에 정확도가 많이 떨어진다. 따라서 안경을 낀 사람의 표정인식을 하려면 데이터셋에 안경 낀 사람이 추가되어야 높은 정확도의 결과를 얻을 수 있을 것이다.

초기에 이 프로그램을 설계할 때는 모델을 resnet50과 vgg16로 사용하였다. 라즈베리파이4의 성능이 비교적 떨어지기 때문에 인식시간이 1분이상이 걸리게 되었다. 높은 정확도를 보여주었지만 사물인터넷에 사용하기에는 경량화가 필요했다.

References

- [1] KOSIS (통계청, 인구총조사), 2020.08.28 [Internet]. Available: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1JC1517&checkFlag=N.
- [2] 이정원. "인공지능 스피커의 의인화와 자기노출이 1인가구의 주관적 행복감과 우울 감소, 삶의 질에 미치는 영향," 석사학위논문, 홍익대학교 산업미술대학원, 서울, 2020.
- [3] 최정우. "공기정화를 위한 아두이노-라즈베리파이 기반의 IoT 플랫폼 개발," 석사학위논문, 성균관대학교 일반대학원, pp.20-25, 서울, 2018.
- [4] 이용준. "임베디드 시스템을 사용한 얼굴 탐색." 석사학위논문, 중앙대학교 대학원, 서울, 2008.
- [5]"kaggle", fer2013 [Internet]. Available: <https://www.kaggle.com/deadskull7/fer2013>.