

# 얼굴표정을 통한 감정 분류 및 음악재생 프로그램

윤경섭\*, 이상원<sup>o</sup>

<sup>o</sup>인하공업전문대학 컴퓨터정보공학과

e-mail: ksyoon@inhac.ac.kr\*, zzxng123@gmail.com<sup>o</sup>

## Music player using emotion classification of facial expressions

Kyung-Seob Yoon\*, SangWon Lee<sup>o</sup>

<sup>o</sup>Dept. of Computer Science, Inha Technical College

### ● 요약 ●

본 논문에서는 감정과 힐링, 머신러닝이라는 주제를 바탕으로 딥러닝을 통한 사용자의 얼굴표정을 인식하고 그 얼굴표정을 기반으로 음악을 재생해주는 얼굴표정 기반의 음악재생 프로그램을 제안한다. 얼굴표정 기반 음악재생 프로그램은 딥러닝 기반의 음악 프로그램으로써, 이미지 인식 분야에서 뛰어난 성능을 보여주고 있는 CNN 모델을 기반으로 얼굴의 표정을 인식할 수 있도록 데이터 학습을 진행하였고, 학습된 모델을 이용하여 웹캠으로부터 사용자의 얼굴표정을 인식하는 것을 통해 사용자의 감정을 추측해낸다. 그 후, 해당 감정에 맞게 감정을 더 증폭시켜줄 수 있도록, 감정과 매칭되는 노래를 재생해주고, 이를 통해, 사용자의 감정이 힐링 및 완화될 수 있도록 도움을 준다.

**키워드:** 감정(emotion), 얼굴표정인식(facial expression recognition), 뮤직플레이어(music player)

### I. Introduction

최근 감성에 대한 관심과 감성의 힐링 및 소비자의 관심사 추천과 같은 인공지능의 활용이 급증하고 있다. 특히, 4차 산업혁명에서 주목받고 있는 머신러닝은 많은 기업들이 작업의 효율성을 증진시키고 비용을 줄이기 위해 많이 적용함으로써, 다양한 분야에서 각광을 받고 있고 현대사회에서 감성은 오늘날의 우리세대는 감성시대라 불리게 할만큼 관심이 많아지고 있다. 감성을 자극하는 요소 중 대표적인 것은 음악으로써, 음악은 다양한 방면에서 사용되어지고 있다. 심리적으로 불안한 사람이나 우울증 등 정신적 질병을 앓고 있는 사람들에게 음악치료 또한 많이 이루어지고 있으며, 뿐만 아니라 보통의 사람들에게도 ‘힐링’의 명목으로 심리적 안정을 취하는데 음악이 빠지지 않고 사용된다. 본 연구에서는 사용자의 기분을 표정에 따라 판별하고 음악을 틀어 줌으로써 감정적 완화에 도움을 줄 수 있도록 음악과 인공지능을 접목한 얼굴표정 인식의 음악재생 프로그램을 개발해보고자 하였다.

본 논문은 다음과 같이 기술된다. 2장에서는 데이터셋의 수집과 전처리에 대해 소개하며, 3장에서는 프로그램의 전체적 설계와 모델학습 및 비교에 대해 보이고 4장에서는 구현 결과를 보인다. 마지막으로 5장에서는 결론을 맺는다.

### II. Preliminaries

머신러닝은 인공지능의 연구 분야 중 하나로, 인간의 학습 능력과 같은 기능을 컴퓨터에서 실현하고자 하는 기술 및 기법이다. 먼저, 본 보고서에서는 감정인식을 위한 머신러닝을 수행하고자 행복, 슬픔, 놀라움 3가지 표정의 데이터셋을 수집하였다. 기존의 공개된 데이터셋의 경우, 서양얼굴 위주의 데이터셋이거나 일본인 여자 또는 적은 표정 수의 특정 환경에서만 운용가능한 한정적인 데이터가 대부분이었다. 따라서, 국내에서 운용 가능한 표정인식 모델을 만들기 위해선 적합한 데이터셋 수집이 필요하다고 보았고 2단계에 걸쳐 Icrawler[1]와 OpenCV[2] 라이브러리를 이용하여 데이터셋 수집을 진행하였다. 1단계에서는, 각 표정마다 Icrawler라는 이미지크롤링 라이브러리를 이용하였으며, 검색 키워드를 다르게하여 Google[3]에서 이미지를 수집하였고 수집된 이미지는 그림 1과 같다.



Fig. 1. Crawling Images

2단계에서는 크롤링한 이미지에 대해 OpenCV라이브러리를 이용하여 얼굴부분을 정사각형으로 잘라 저장하였다. 얼굴영역을 추출한 이미지는 그림 2와 같다.



Fig. 2. Extract face region

이 과정에서, 어느정도의 필요없는 이미지는 제거되었으나, 여전히 관계없는 이미지들이 많았고 이미지를 전체적으로 눈으로 보며 필터링 함으로써, 각 표정당 500개씩의 데이터를 생성하였다. 그 후, 생성된 데이터에 대해 학습하기에 알맞은 형태로 바꾸기 위해서 그림 3과 같이 32x32픽셀의 이미지로 Resize를 진행한 후, 회색조로 바꾸어 이미지의 전처리를 수행하였다.

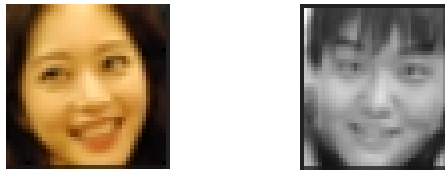


Fig. 3. data preprocessing

이 과정들을 통해 표정별 500개씩 데이터를 구축하였으며, 각 표정별 데이터 500개씩에 대해 후미의 50개를 테스트 데이터로 나머지 450개의 데이터를 트레이닝 데이터로 사용하였다.

### III. Design

#### 3.1. 표정인식 음악프로그램 설계

본 연구에서 개발하고자하는 표정인식 음악프로그램은 그림4와 같이 왼쪽부분의 음악재생프로그램과 오른쪽부분의 웹캠을 통한 얼굴 표정인식부분을 각각 나누어 얼굴표정인식 기능을 사용하지 않더라도 그자체로 뮤직플레이어로서의 기능을 수행할 수 있도록 따로 구현하고 합치도록 GUI를 설계하였다.

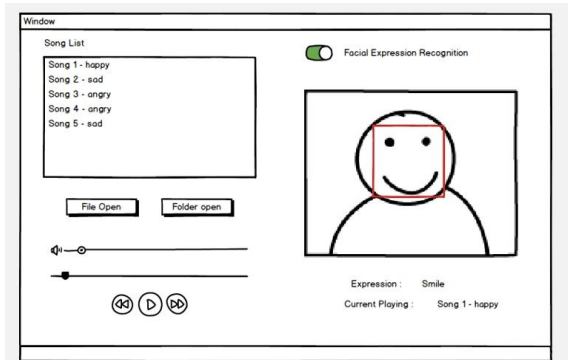


Fig. 4. GUI Design

기능면에서는 얼굴표정인식부터 음악이 재생되기까지 그림5와 같은 순서로 진행되도록 설계하였다.



Fig. 5. Algorithm of Music play using Facial expression recognition

#### 3.2. CNN 학습 및 NN과의 성능 비교

얼굴표정 기반 음악재생 프로그램을 만들기 위해 수집한데이터를 기반으로 이미지 분류에 있어서 강점을 보이고 있는 CNN[4]을 이용하여 얼굴분류모델을 학습하였다.

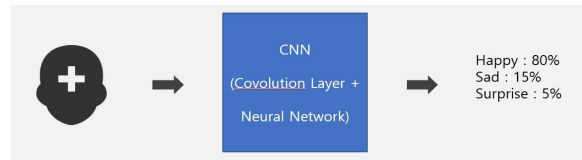


Fig. 6. CNN

먼저, CNN(Convolution Neural Network)과 NN(Neural Network)와의 차이점으로 꼽는다면 CNN은 답러닝으로써, 일반적인 머신러닝과는 다르게 자체에서 특징을 추출하고 학습한다는 점이다. 따라서, 이미지에 대해 특징 학습을 위해 Convolution 계층과 Pool 계층이 들어가 있는 것이고 이 부분을 구성해야했다. CNN의 대표적인 예제인, mnist 글자인식[5]을 참조하여 2개층으로 구현하였으며, NN의 학습결과와 CNN의 학습결과를 비교하여 보았다.

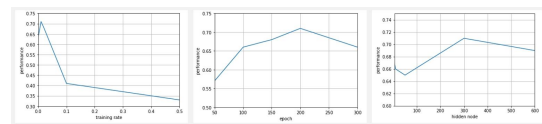


Fig. 7. NN Performance (Training rate, Epoch, Hidden nodes)

먼저, NN만으로 학습률, 에포크, 히든 노드수의 변화를 쥐가면서 성능을 체크해 보았다. 그 결과, 학습률이 0.01이고 에포크가 200, 히든 노드수가 300일 때에 정확도가 약 71%로 가장 높은 성능을 보임을 확인할 수 있었다.

Table 1. CNN Accuracy

테스트 순번	정확도
1	0.880
2	0.889
3	0.898
4	0.907
5	0.889
6	0.889
7	0.907
8	0.889
9	0.907
10	0.917

CNN에서는 학습에 있어서 step을 10000으로 loss를 측정한 결과 0.11까지 떨어졌으며, 테스트 데이터셋을 가지고 표1과 같이 정확도를 10번 측정하여 평균 내본 결과 89.72%로 일반적으로 NN를 사용할 때보다 약20%정도 성능이 향상된 것을 확인할 수 있었다.

#### IV. Implementation

얼굴표정 기반 음악재생 프로그램을 구현하기 위해, 뮤직플레이어를 부분과 학습된 모델을 바탕으로한 얼굴인식 부분을 구현하였다. GUI는 PyQt5[6]를 이용하고 Python을 기반으로 프로그램을 제작하였다.



Fig. 8. Program Result

그림8은 구현된 최종적으로 구현된 화면을 보여주며 왼쪽 부분은 음악재생리스트이다.



Fig. 9. Music List

메뉴탭에서 음악을 지정하고 해당 곡의 분위기를 지정해줌으로써

그림9와 같이 음악이 추가되어지며, 얼굴인식 없이도 하나의 뮤직플레이어으로써, 사용되어질 수 있도록 하였다.

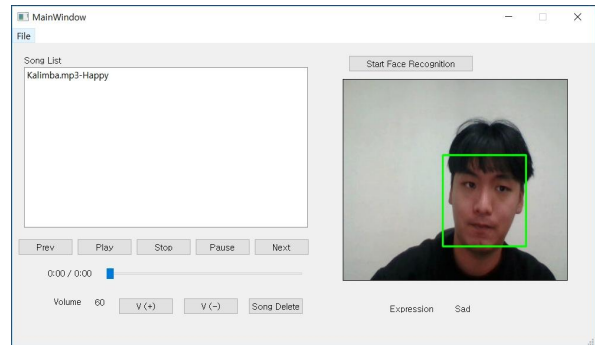


Fig. 10. Facial Expression Recognition

그림10은 얼굴표정인식을 통한 음악재생으로써 행복, 슬픔, 놀라움 중 한 표정을 5초간 유지하면 해당 표정과 매칭되는 곡이 자동으로 재생되어지도록 하였다.

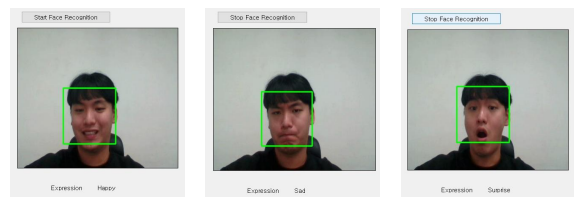


Fig. 11. Each Facial Expression Recognition

각 얼굴표정인식은 그림11과 같다.

#### V. Conclusions

이 연구를 통해 사용자의 얼굴표정을 읽고 감정을 분류하여 해당감정에 따라 감정과 매칭되는 음악을 틀어줌으로써, 사용자의 감정이 완화되고 심리적으로 안정을 찾을 수 있도록 도와줄 수 있을거라 생각한다. 향후 계획으로는 음악 추가시 현재에는 사용자가 수동으로 해당곡의 분위기를 지정하도록 되어있는데, 이 부분을 FFT[7](시간에 따른 신호의 변화, 주파수)알고리즘을 이용하여 각 분위기별 노래의 주파수 변화를 특징으로 활용하고, 머신러닝을 통해 학습시킴으로써 분위기를 자동 지정해줄 수 있도록 연구해볼 계획이다.

#### REFERENCES

[1] Icrawler, Image Crawling Library, <https://pypi.org/project/icrawler/>  
 [2] OpenCV, Open Source ComputerVision Library, <https://opencv.org/>

- [3] Google Image, Image Research Site, [https://www.google.com/imghp?hl=ko\\_\\_](https://www.google.com/imghp?hl=ko__)
- [4] CNN, Convolution Neural Network is a class of deep neural networks, most commonly applied to analyzing visual imagery, [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)
- [5] Mnist, The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems, [https://en.wikipedia.org/wiki/MNIST\\_database](https://en.wikipedia.org/wiki/MNIST_database)
- [6] PyQt5, Library for configuring the GUI in Python, <https://pypi.org/project/PyQt5/>
- [7] FFT, algorithm that computes the discrete Fourier transform (DFT) of a sequence, or its inverse (IDFT), [https://en.wikipedia.org/wiki/Fast\\_Fourier\\_transform](https://en.wikipedia.org/wiki/Fast_Fourier_transform)