

이미지 분위기 분류에 기반한 동영상 자동 생성

조동희, 남용욱, 이현창, 김용혁

광운대학교

whehd16@naver.com, mitssimvz@gmail.com, qzeczxad@naver.com, yhdfly@kw.ac.kr

Automatic Video Generation Based on Image Mood Classification

Dong-Hee Cho, Yong-Wook Nam, Hyun-Chang Lee, Yong-Hyuk Kim

Kwangwoon University

요약

머신러닝을 활용한 이미지 분류는 단순 사물을 넘어서 사람의 감정과 같은 추상적이고 주관적인 개념에도 적용되고 있다. 이 중에서도 합성곱 신경망을 통한 이미지의 감정 분류 연구가 더욱 활성화되고 있다. 그럼에도 다양한 멀티미디어들을 머신러닝 알고리즘으로 분석하고 이를 의미있는 결과로 재구성하기는 매우 복잡하고 까다롭다. 본 연구에서는 기존 연구를 개선시켜 음악 데이터를 다층퍼셉트론 모델을 통해 분류된 이미지와 결합한 동영상과 파이프라인의 다양한 라이브러리를 통해 자동으로 생성하였다. 이를 통해 특정 분위기로 분류된 이미지들과 이에 어울리는 음악을 매칭시켜 유의미한 새로운 멀티미디어를 자동으로 생성할 수 있었다.

1. 서론

최근 들어 머신러닝을 활용한 이미지 분류는 단순 사물을 넘어서, 사람의 감정과 같은 추상적이고 주관적인 개념에도 적용되고 있다. Kim[1]의 연구에서는 머신러닝에 감정 분류 기법을 결합한 모델을 사용하여 영상의 감성을 분석하고자 했다.

다양한 기법 중에서도 합성곱 신경망(convolutional neural network; CNN)을 활용한 이미지 분류는 머신러닝 분야에서 많은 이론의 적립과 유용한 성과[2]를 거두고 있으며, 다양한 응용 분야에도 적용되고 있다. Cho 등[3]의 연구에서는 심층 합성곱 신경망을 활용하여 영상의 분위기 분류를 진행했다. 음성과 이미지 등과 같은 다양한 정보를 다루는 멀티미디어들을 머신러닝 알고리즘으로 분석하고, 이를 의미 있는 결과로 재구성하는 것은 매우 복잡하고 까다롭다. 기존 연구 [3]에서는 머신러닝 기법을 통해 이미지의 분위기를 8가지로 분류하고 이에 해당하는 음악 데이터와 결합한 결과를 개별적으로 처리하여 동영상을 완전히 자동적으로 구성하지 못했다. 본 논문에서는 이를 보완하고 개선하여 이미지와 음악을 하나의 동영상을 자동적으로 생성하는 방안을 제시한다.

본 논문의 구성은 다음과 같다. 2 절에서는 분위기 모델 및 학습 모델에 사용된 머신러닝 기법과 개선된 동영상 자동 생성 기법을 소개한다. 3 절에서는 결과 및 향후 연구를 소개한다.

2. 분위기 분류 모델 및 동영상 자동 생성 기법

감사의 글 : 본 논문은 2019년도 교육부의 대학혁신지원사업 사업비를 지원받아 수행된 연구임.

2.1 분위기 모델

본 연구에서는 Cho 등[3]의 8 가지 분위기 분류 모델을 사용한다. 이는 인간의 감정을 포괄하면서 실제로 사람이 느낄 수 있는 분위기까지 고려한 것이다. Fig. 1은 이에 해당되는 분위기 분류 모델이다. 이웃한 분위기들은 서로 다르지만 비슷한 뉘앙스를 가진다는 점에서 색의 스펙트럼과 같은 자연스러운 분위기 분포를 표현했다.

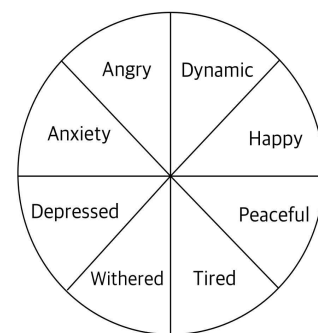


Fig. 1. Mood classification model[3]

2.2 다층 퍼셉트론

이미지의 분위기 분류는 머신러닝의 심층 합성곱 모델을 사용했다. 연구에 사용된 모델은 여러 개의 합성곱과 정류된 선형 유닛(rectified linear unit; ReLU) 활성화 함수 그리고 덴스(dense)층이 반복적으로 구성되어 있는 다층 퍼셉트론(multilayer perceptron)[4]을 사용한다. 모델의 손실과 과적합을 줄이기 위해 각각 소프트맥스-교차 엔트로피 손실 함수(softmax - cross entropy loss function)와 드롭

아웃(dropout)을 사용하였다. 학습에 사용된 에포크(Epoch) 수와 배치 크기(Batch Size)는 각각 1,000, 64이다. Fig.2 는 다층 퍼셉트론의 내부 구조 예시이다.

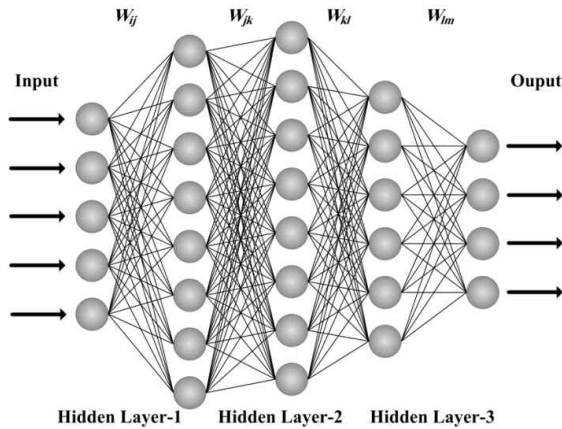


Fig.2. Example of a multilayer perceptron

2.3 동영상 자동 생성

기존 연구에선 이미지와 음악 데이터를 사용자가 직접 동기화 및 편집하여 동영상을 생성하는 방식으로, 각각을 분리해서 처리하는 구성으로 되어 있었다. 이는 음악과 이미지가 결합한 동영상의 완전한 자동 생성으로 보기 힘들다. 본 연구에서 제안하는 방식은 파이썬 (python)의 OpenCV(open source computer vision)[4], 필로우 (pillow), 파이게임(pygame)와 같은 이미지, 오디오 처리가 가능한 라이브러리를 활용하는 것이다. OpenCV는 실시간 컴퓨터 비전을 목적으로 한 라이브러리로 이미지 프로세싱에 중점을 둔 라이브러리이다. 필로우는 PIL(python imaging library)과의 코드 호환성이 뛰어나고 이미지 처리를 더욱 쉽게 할 수 있는 라이브러리이다. 파이게임은 GUI 기반의 다양한 동작을 간편하게 처리할 수 있는 라이브러리이다.

특히 동영상 제작에 있어서 페이드 인(fade-in), 페이드 아웃(fade-out) 효과를 중점적으로 부여했다. 이미지 출력과 음악 출력에 페이드 인 시간과 페이드 아웃 시간을 부여하여, 음악의 재생 볼륨이 이미지가 점차 생성되거나 사라지는 모습에 비례하도록 재생하였다. 이를 통해 둘을 하나의 데이터처럼 만들고 다음 데이터와의 자연스러운 연결을 통해 끊김 없는 동영상을 자동 생성할 수 있다. Fig.3은 생성한 실제 이미지와 음악이 결합 되어 재생되고 있는 동영상에서 캡처한 것이고 내부에서 이미지의 시간에 따른 페이드 인, 페이드 아웃 효과를 확인할 수 있다. 연구를 통해 생성된 동영상은 이 링크[2]에서 확인할 수 있다. 그리고 동영상 제작에 사용된 32장의 일반적인 이미지는 파이썬을 통한 웹 크롤링(web crawling)[5] 방식을 통해 직접 추출했으며, 모델을 학습시킬 때 사용했던 데이터로부터 독립적이다.

3. 결과 및 향후 연구

본 연구를 통해 다층퍼셉트론으로 구성되는 심층 합성곱 신경망 모델을 활용하여 이미지를 8가지의 분위기로 분류하고 이에 맞는 음악

데이터를 결합하여 다양한 효과가 포함된 자연스러운 동영상을 자동으로 생성할 수 있었다. 본 연구를 통해 기존 연구보다 더욱 개선된 자동화 방식으로 동영상을 생성하는 결과를 얻었다. 향후 연구에선, 실시간으로 동영상의 분위기를 분석하여 음악을 부드럽게 결합해주는 연구와 더욱 구체화 된 감정 분류를 통한 멀티미디어와의 융합을 보여주는 연구가 가능할 것이다.

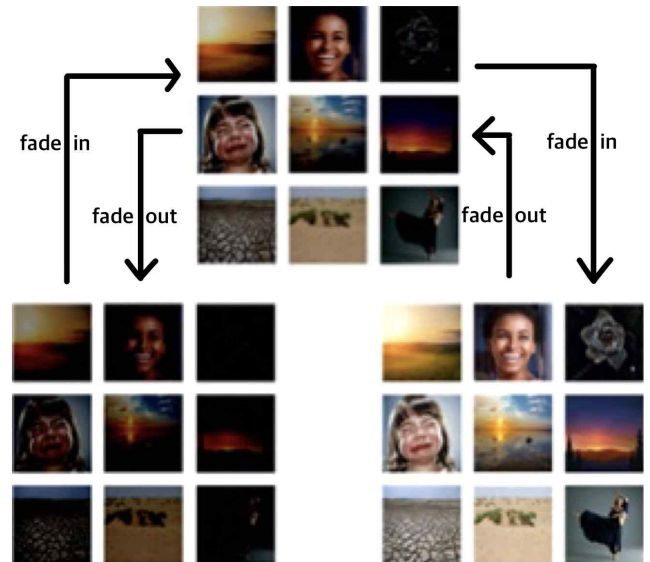


Fig.3 A capture of images in our generated video using fade-in and fade-out

참 고 문 헌

- [1] S. H. Kim. (2016). Sentiment classification for videos using deep learning algorithms. Master dissertation. Seoul University, Seoul.
- [2] A. Sharif Razavian, H. Azizpour, J. Sullivan, & S. Carlsson (2014). CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 806-813).
- [3] D. H. Cho, Y. W. Nam, H. C. Lee, Y. H. Kim. (2019). Image Mood Classification Using Deep CNN and Its Application to Automatic Video Generation. Journal of the Korea Convergence Society, 10(9), 23-29.
- [4] G. Bradski & A. Kaehler (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.". 2008.
- [5] S. Chakrabarti, B. E. Dom & M. H. van den Berg (2002). U.S. Patent No. 6,418,433. Washington, DC: U.S. Patent and Trademark Office.

2) <https://youtu.be/u4cl-GkyiwA>