컴퓨터비전과 자연어 처리를 활용한 표정 인식 디지털 미디어아트 구현에 관한 연구

한재상, 강윤서, 권채연, 허원제, 김영종
[1]숭실대학교 소프트웨어학부
✉숭실대학교 소프트웨어학부
hhncn4471@soongsil.ac.kr, ysys1130@gmail.com
rnjscodus130@naver.com, heowj03145@naver.com

A Study on the Implement of Digital Media Art Using Computer Vision and Natural Language Processing

Han Jae-sang, Kang Yoon-seo,
Kwon Chae Yeon, Heo Won-Jae, Youngjong Kim✉
*School of Software, Soongsil University
✉School of Software, Soongsil University

요 약

본 논문은 자연어 처리와 Face Detection 기술을 활용하여 사용자의 현재 감정 상태에 따른 디지털 미디어 아트를 구현하는 방법을 기술한다. 사용자의 현재 표정과 감정을 분석하고, 감정에 따라 다양한 시각 효과 및 다른 성격의 텍스트를 화면에 송출하는 방식으로 사용자와 상호작용한다.

1. 서론

새로운 기술과 사회 환경은 개인 간 소통의 방식을 변화시킨다. 이러한 소통 방식의 변화는 첨단기술과 융합하여 현대 디지털 미디어 아트의 오브제로 활용되기도 한다. 현재 코로나19 대유행은 감염병과 이로 인한 공동체 구성원의 사망으로 대표되는 공공보건의 위기이며, 급속한 사회 환경 변화와 소통 방식의 전환을 불러왔다. 이 위기는 단순히 감염자의 확산으로 인한 의료체계의 붕괴에 그치지 않고 일상생활에서의 우울과 불안, 대인관계 단절과 사회적 소통 미흡 등 개개인의 삶에 부정적인 영향을 미쳤다. 코로나19 방역으로 인해 경제활동과 일상생활에 많은 제약이 뒤따르면서 우울감을 호소하고 불안장애를 보이는 사람들이 OECD 주요국(15개국) 평균 3배 이상 증가하였다. [우울 수준(6.8%→21.8%) [그림 1] 참조].

자료: OECD, Tackling the mental health impact of the COVID-19 crisis: An integrated, whole-of-society response, 2021. p.4.

최근 코로나19가 정신건강에 미친 영향에 대한 OECD 연구(2021)에 따르면, 코로나19 이후 우울 수준의 경우 한국은 OECD 평균(21.8%)을 상회하여 비교 대상 8개국 표시 중 1위(36.8%)를 차지하고 있다. 일상생활에서 '관계'와 '소통'은 국민 개인의 행복과 안녕에 미치는 영향이 지대하다. 하지만 코로나19 대유행은 일상생활을 제약함으로써 대인관계를 어렵게 하는 동시에 고립으로 인해 정신건강에 부정적인 영향을 미친다[1]. 이러한 사회 분위기 속에서 사람들은 감정을 표현하는 방법을 망각하고 타인과 친밀한 관계를 맺는 것에 어려움을 겪고 있다.

본 팀은 지난 2년간 도외시되었던 비언어적 소통의 복원과, 힘들고 지쳤던 사람들을 위로하기 위한 목적으로 본 프로젝트를 시작하였다.

2. 프로젝트 개요

본 프로젝트는 사용자의 표정에 따라 다양한 시각 효과를 연출하는 디지털 미디어 아트로서, 사용자에게 표정을 통해 디지털 세상과 소통하는 경험을 제공한다. 이를 위하여 컴퓨터 비전의 주요 연구 분야 중 하나인 Face Detection 기술과 문장별 감정을 자연어 처리로 수치화한 텍스트 데이터를 활용한다.

전체적인 소프트웨어 인터페이스 구조는 다음 그림과 같다.

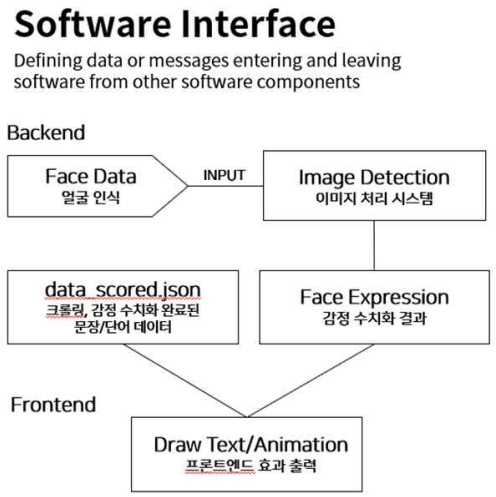

그림 2 소프트웨어 인터페이스 개관

Backend 에서의 주요 시스템은 Face Recognition, Text Data Sampling으로 나뉜다.

Face Recognition 시스템은 사용자의 얼굴을 인식하여 표정에 따른 감정을 수치화하는 기능을 담당한다. 주요 Input 값은 사용자의 실시간 영상 데이터로서, 적절한 형식으로 Face Data가 입력된다면 n프레임 단위로 장면을 추출하여 Image Detection 프로세스를 수행한다. 이 과정에서 Face Detection Algorithm에 따라 사용자의 얼굴을 인식하며 눈, 코, 입 위치 등의 Facial landmarks와 눈 사이의 거리, 이마에서 턱까지의 거리 등의 Key factors를 바탕으로 사람의 얼굴을 인식한다. 얼굴 인식을 통해 도출된 Vertex 정보는 Face Expression 감지 모듈에 전달되어 표정에 따른 감정을 인식하는 기능을 수행한다. 이렇게 만들어진 Face Expressions 객체는 Frontend에 전달된다.

Text Data Sampling에서는 커뮤니티 게시물, 문학작품, 노래 가사 등 디지털 미디어에서 크롤링해온 단어 및 문장을 감정에 따라 수치를 매겨 전처리하는 기능을 담당한다. 이는 Frontend 단에서 감정에 따라 다양한 시각 효과를 연출할 때, 각기 다른 성격의 텍스트를 화면에 송출하기 위해 필요하다. 각 문장별 Positive, Negative 정도를 판단하기 위해 자연어 처리 기술이 활용되며, 수치화된 데이터는 Frontend에 json 형식으로 전달된다.

Frontend 단에서는 전달받은 Face Expressions 값에 따라 사용자의 현재 감정 상태를 정의하고, 감정에 따른 시각 효과를 연출한다. 또한 샘플링된 텍스트 정보를 이용해 텍스트 애니메이션을 출력한다.

3. 프로젝트 기능 구현

3.1 Text Parsing & Analyze Sentiment

자연어 처리(NLP: Natural Language Processing)는 사람의 언어인 자연어를 기계인 컴퓨터가 이해할 수 있도록 하는 처리과정 및 기술이다. 본 연구에서 사용한 '구글 자연어 처리 API(Google Natural Language API)'는 오픈소스 API로 총 5개의 Method를 제공한다[2]. 이들은 감성 분석(Sentiment Analysis), 항목 분석(Entity Analysis), 항목 감성 분석(Entity Sentiment Analysis), 구문 분석(Syntax Analysis), 콘텐츠 분류(Content Classification)이다. 본 연구에서는 이들 중 감성 분석을 활용하며 이것은 영어, 한국어를 포함한 다국어를 지원한다.

감성 분석결과는 문서 전체의 전반적인 감정값과 해당 문서에서 추출된 문장의 목록 및 문장 단위의 감정 값을 포함한다. 기계의 관점에서 바라보는 감정은 두 가지 척도로 치환된다. 긍정과 부정의 양극성과 이에 따른 감정의 값으로 분류되는 score(-1에서 1 사이의 값), 그리고 감정의 강도인 magnitude(0에서 무한대까지의 값) 두 항목이 이에 해당한다. Magnitude는 score의 절댓값으로 표현되며 값이 클수록 감정의 강도가 세지는 것으로 해석할 수 있다. score와 달리 magnitude는 정규화되지 않는다. 텍스트 내의 각 감정 표현(긍정적 및 부정적 모두)이 텍스트 magnitude에 반영되며, 따라서 긴 텍스트 블록일수록 값이 더 커진다. magnitude 값의 총합을 통해 문서 전체의 감정의 강도를 측정할 수 있고 문서 전체의 정서 성향(긍정/부정)은 각 문장별 score값의 평균으로 해석할 수 있다. 0을 중립적인 값으로 양극단의 값(-1 ~ 1)으로의 거리를 통해 긍정적이거나 부정적인 감정의 성향을 파악할 수 있다. 그리고 감정의 강도를 0에서부터 커진 양에 따라 판단할 수 있다[2].

중립적 점수(0.0 근처)의 문서는 감정적 콘텐츠가 많거나 복합적 감정이 포함되어 긍정적 값과 부정적 값이 서로를 상쇄시키는 문서일 수 있다. 일반적으로 magnitude 값을 사용하여 이 두 가지 경우를 구분할 수 있다. 문서가 실제로 감정 중립적이라면 magnitude 값이 낮지만, 복합적 감정이 포함된 문서는 magnitude 값이 높게 측정된다 [3].

아래 [표 1]에서는 하나의 문서전체의 감성분석 결과를 기반으로 그 값을 해석한 결과에 대해 보여준다[4]. 해당 값들이 감정의 성향이나 그 성향의 정도, 감정의 강도를 수치상으로 출력하여 주지만 그것이 어떤 세분된 감정을 나타내는지는 알 수 없다.

sentiment	value
very positive	“score”: 0.8, “magnitude”: 3.0
positive	“score”: -0.6, “magnitude”: 4.0
neutral	“score”: 0.1, “magnitude”: 0.0
mix	“score”: 0.0, “magnitude”: 4.0

표 2 구글 자연어 처리 감정분석 API 값의 해석

3.2 Face Detection 구현

Face Recognition 기술은 금융, 보안 등 다양한 응용 분야에 적용되어 왔으며, 그 적용 범위가 점차 확대되고 있다. 초기의 얼굴 인식 기술은 이미지 센서 및 처리 시스템의 SW/HW 리소스 제약으로, 높은 실패율과 낮은 처리 효율을 보여 적용범위가 넓지 않았다. 그러나 현대에 이르러 표정, 피부색, 조명의 다양성 문제 등을 해결할 수 있는 다양한 알고리즘과 모델이 제안되었다.[5] 본 연구에서 다루는 실시간 디지털 미디어 아트의 구현을 위해서는 처리가 빠르고 보다 경량화된 얼굴 인식 기술이 필요하였다. 따라서 구현에 앞서 Python 환경에서 MTCNN, SSD MobileNetV1 모델의 얼굴 인식 기술 성능 측정을 진행하였다.

그 결과, MTCNN의 속도보다 SSD MobileNetV1으로 처리를 진행하였을 때 3배 이상의 속도 차이를 보였다. 실시간으로 표정을 인식하고 인식 결과에 따라 효과를 출력해야 하는 본 프로젝트의 특성을 고려하여, SSD MobileNetV1 방식의 얼굴인식 시스템을 적용하였다.

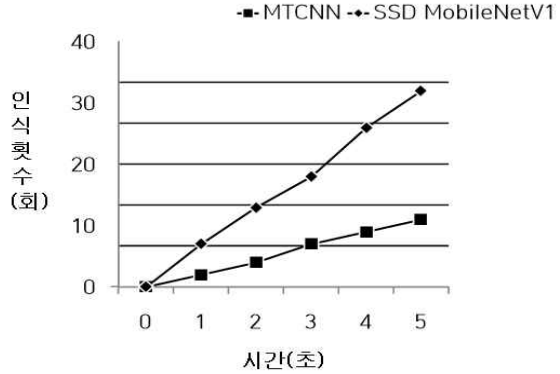

그림 3 MTCNN와 SSD MobileNetV1의 속도 측정

3.3 Face Expression 감지 구현

본 프로젝트에서 사용자는 각기 다른 표정으로 디바이스와 실시간 상호작용한다. 그러므로 정확하고 빠르게 표정 감정(Face Expression)을 읽어내어 서비스에 전달할 수 있어야 한다. 이를 위하여 크게 세 가지 과정으로 나누어 표정 감지를 구현한다. 첫 번째 단계는 정점(Vertex) 정보 추출이다. 사용자 얼굴의 윤곽, 눈, 코, 입 등의 주요 특징을 감지하여 Vertex 좌표 객체로 변환하는 과정이 이에 포함된다. 두 번째 단계는 크기 재조정(resize)다. 앞서 추출한 Vertex 정보를 화면에 출력할 캔버스 크기에 따라 비율을 재조정하여야 올바른 출력 결과를 얻을 수 있다. 마지막 단계는 얼굴 표정 인식(Detect Face Expressions)이다. 이 단계에서 파라미터로 Vertex 정보가 전달되면, 최종적으로 사용자의 현재 감정 상태를 백분율로 수치화하여 객체를 리턴한다. 감정 상태와 조건은 다음 테이블과 같다.

expressions	condition
neutral	default
happy	expressions.happy >= 0.9
fearful	expressions.fearful >= 0.9
surprised	expressions.surprised >= 0.9
angry	expressions.angry >= 0.7
sad	expressions.sad >= 0.7

표 3 Face expressions / condition

happy, fearful, surprised 감정의 경우 0.9 이상의 인식률을 보여야 현재 state로 확정할 수 있도록 설정한다. 이는 표정 인식에서 가장 자주 인식되는 감정이 행복, 두려움, 놀람의 세 감정이기 때문으로, 혼합된 감정을 엄격하게 분류하기 위하여 높은 기준치를 설정한 것이다[6]. 실제로도 테스트 결과, surprised와 angry, fearful과 sad는 혼합되어 나오는 경우가 많았고, 대체로 surprised와 fearful이 우세하게 인식되었다. 기준치를 다르게 설정하여 이와 같은 문제를 완화할 수 있었다. 한편 인식률이 상대적으로 낮은 angry, sad 감정은 0.7 이상의 인식률을 보일 때 state를 확정할 수 있도록 한다. 감정별로 기준치를 설정한다면 복합적인 감정이 균형있게 인식되어 다양한 시각 효과를 연출할 수 있다.

만약 어떠한 다른 condition도 만족하지 않는다면 기본 상태는 default 값인 neutral로 설정된다.

4. 결론

본 논문은 사용자의 표정에 따라 다양한 시각 효과를 연출하는 디지털 미디어 아트의 구현 방법을 제시하였다. 컴퓨터 비전 기술로 사용자의 얼굴 표정을 인식해 감정을 분석한 후. 도출된 감정 상태에 따라 샘플링된 글귀가 포함된 타이포그래피, 애니메이션 등 여러 시각적 효과를 제공하는 프로세스를 제안하였다. 실시간으로 이미지를 인식하여야 하는 본 프로젝트의 특성을 고려하여 실시간 이미지 인식 분야에서 뛰어난 성능을 보여주고 있는 SSD MobileNetV1 모델을 기반으로 얼굴 표정을 인식할 수 있도록 설계하였고, 검증 결과 MTCNN 모델에 비하여 약 3배정도 빠른 속도를 보임을 확인할 수 있었다. 웹캠에 인식된 사용자의 얼굴 표정을 토대로 사용자의 감정을 추측하는 과정에서 특정 감정이 편향되어 인식되는 문제가 발생하였는데, 이는 가중치 및 기준값을 개별적으로 설정하여 문제를 해결할 수 있었다. 이러한 과정을 통해 인식된 사용자의 감정 상태는 시각적 효과로 표현된다. 우선 여러 미디어 매체에서 수집한 텍스트들을 Google Natural Language API를 통해 감정별로 분류하였다. 이후 표정에서 추출한 감정과 일치하는 텍스트, 다양한 시각적 효과를 화면에 송출함으로써 해당 감정을 좋은 방향으로 증폭시키고 불편한 감정들은 완화될 수 있도록 유도하였다. 본 프로젝트는 접근성이 높은 공공장소에 설치하여 사람들에게 심리적 위안과 공감, 격려의 메시지를 전하는 데에 활용하고자 한다. 기술적 요소를 예술과 융합하여 공공 가치와 메시지를 전달한다는 점에서, 인터랙티브 미디어아트를 구현하는 데에 본 연구가 참고가 될 수 있기를 바란다. 다만, 본 프로젝트는 얼굴 표정만을 기반으로 사람의 감정을 판단하기 때문에 정확도 측면에서 한계가 존재한다. 따라서 이를 보완하기 위해 비언어적 표현 및 반언어적 표현까지 고려할 수 있도록 후속 연구가 필요하다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2018-0-00209)

참고문헌

[1] 국회입법조사처, ['코로나19 대유행'이 가져온 정신건강 위기와 대응 정책과제], 이슈와 논점 제1880호, 2021

[2] https://cloud.google.com/natural-language/docs/basics?hl=ko

[3] 김혜란, [가사 텍스트의 감성분석에 기반 한 음악 시각화 콘텐츠 개발], 세종대학교, 2020, p.91-93

[4] https://cloud.google.com/natural-language/docs/basics?hl=ko%2C%2 0%202019.10.3.#interpreting_sentiment_analysis_values

[5] Di Lu, Limin Yan, [Face Detection and Recognition Algorithm in Digital Image Based on Computer Vision Sensor], Journal of Sensors, vol. 2021, Article ID 4796768, 16 pages, 2021.

[6] Fraser W. Smith, and Stephanie Rossit, [Identifying and detecting facial expressions of emotion in peripheral vision], plos one, 13(5): e0197160. 2018.

저자소개

한 재 상 (Jaesang Han)

팀에서 컴퓨터 비전을 활용한 Backend 개발을 담당하고 있으며, 현재 숭실대학교 소프트웨어학부에 재학중이다. 관심 분야는 SW공학 및 오픈소스 창업이다.

강 윤 서(Yunseo Kang)

팀에서 텍스트 데이터를 처리하는 Backend 개발을 맡고 있으며, 현재 숭실대학교 소프트웨어학부에 재학 중이다. 관심 분야는 보안 프로그래밍과 백엔드 개발이다.

권 채 연 (Chaeyeon Kwon)

팀에서 텍스트 애니메이션 Frontend 개발과 설계를 담당하고 있으며, 현재 숭실대학교 소프트웨어학부에 재학중이다. 관심분야는 Web 및 App 서비스 개발이다.

허 원 제 (Wonje Heo)

팀에서 프론트엔드 설계를 담당하고 있으며 현재 숭실대학교 소프트웨어학부에 재학 중이다. 관심분야는 인공지능 Open API 개발과 빅데이터 분석이다.

김 영 종 (Youngjong Kim)

한글과컴퓨터, 씽크프리 연구원, 시큐어소프트 기술이사, 하우리 대표이사, 오픈소스커뮤니티연구소 소장 등을 역임하였으며, 현재 숭실대학교 소프트웨어학부 교수로 재직중이다. 관심분야는 보안 및 클라우드컴퓨팅, 블록체인이며 오픈소스기반의 클라우드 및 블록체인 프로젝트를 수행하고 있다.