

좌표계산을 통해 동영상의 안면 특징점 분석을 중심으로 한 웹 기반 발표 태도 교정 프로그램 개발

Development of a Web-based Presentation Attitude Correction Program Centered on Analyzing Facial Features of Videos through Coordinate Calculation

권기현, 안수호, 박찬정
제주대학교 컴퓨터교육과

Kihyeon Kwon(rlgus5126@naver.com), Suho An(ahnshoz@icloud.com),
Chan Jung Park(cjpark@jejunu.ac.kr)

요약

학생들의 취업을 위한 면접 발표와 회사에서의 프로젝트 결과 발표 등과 같은 형식적인 발표 태도가 개선되려면 동료나 교수자의 관찰에 의한 방법 이외에 자동화된 방법은 드물다. 기존 연구에 따르면, 발표자의 안정적인 발화와 시선 처리가 발표에서의 전달력에 영향을 미친다고 한다. 또한, 본인 발표에 대한 적절한 피드백이 발표자의 발표 역량을 높이는 효과가 있다는 연구도 있다. 본 연구에서는 이와 같은 교정의 긍정적 측면을 고려하여 대학생들의 잘못된 발표 습관과 태도를 동영상의 안면 분석을 통해 지능적으로 교정해 주는 프로그램을 개발하고 성능을 분석하였다. 개발하는 프로그램은 웹 기반으로 군말 사용 여부를 확인하고 안면 인식과 발표 내용 텍스트화를 통해 개발되었다. 이를 위해 군말 분류 인공지능 모델을 개발하였고, 동영상 객체 추출 후, 좌표에 기반으로 얼굴 특징점을 인식하였다. 이후 4,000개 안면 데이터를 이용해 Teachable Machine에서 안면 인식한 경우와 본 연구의 알고리즘 성능을 비교·분석하였다. 프로그램을 이용해 발표 태도를 자기 스스로 교정하여 발표자들에게 도움을 준다.

■ 중심어 : | 발표 태도 | 얼굴 인식 | 동영상 분석 | 음성 인식 | 특징점 추출 |

Abstract

In order to improve formal presentation attitudes such as presentation of job interviews and presentation of project results at the company, there are few automated methods other than observation by colleagues or professors. In previous studies, it was reported that the speaker's stable speech and gaze processing affect the delivery power in the presentation. Also, there are studies that show that proper feedback on one's presentation has the effect of increasing the presenter's ability to present. In this paper, considering the positive aspects of correction, we developed a program that intelligently corrects the wrong presentation habits and attitudes of college students through facial analysis of videos and analyzed the proposed program's performance. The proposed program was developed through web-based verification of the use of redundant words and facial recognition and textualization of the presentation contents. To this end, an artificial intelligence model for classification was developed, and after extracting the video object, facial feature points were recognized based on the coordinates. Then, using 4000 facial data, the performance of the algorithm in this paper was compared and analyzed with the case of facial recognition using a Teachable Machine. Use the program to help presenters by correcting their presentation attitude.

■ keyword : | Presentation Attitude | Face Recognition | Video Analysis | Audio Analysis | Feature Extraction |

* 본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음.(2018-0-01863)

* 본 논문은 한국콘텐츠학회 2021 종합학술대회 우수논문입니다.

접수일자 : 2022년 01월 06일

심사완료일 : 2022년 01월 26일

수정일자 : 2022년 01월 26일

교신저자 : 박찬정, e-mail : cjpark@jejunu.ac.kr

I. 서론

최근 2022 개정 교육 과정의 총론이 발표되었고, 새로 나온 교육 과정을 위한 학습자들의 핵심 역량 중 협력적 의사소통이 포함되어 있다[1]. 더불어 최근 학습자 중심의 교수-학습 방법이 중요시되면서, 토론이나 발표 위주의 수업이 학교 현장에서 많이 이루어지고 있다[2]. 학교 현장이나 대학 기관에서는 취업이나 사회생활 등을 위해 공식적으로 말하기 유형인 '발표'가 많이 이루어지고 있다. 발표는 내용의 단순한 전달이 아니라 문제의 핵심을 분석하여, 관련 자료를 수집하고 이를 체계적으로 정리하여, 제한된 시간에 자료와 함께 설명이나 설득을 하는 의사소통 행위이다. 따라서, 학습 활동으로 중요하다[3].

발표 중간에 얼굴이나 머리카락을 만진다든지, 다리를 떤다든지, 청중이 들리지 않을 정도로 작은 목소리로 말한다든지, 외위 입듯이 단조로운 어조로 말한다든지, 같은 말을 반복한다든지, 청중과 시선을 잘 마주치지 못하고 책상에 있는 발표문을 응시하며 말을 하는 등 잘못된 습관을 지니고 있다. 학생들은 발표 연습과 더불어 발표에 대한 피드백을 통해 자신의 잘못된 발표 습관을 고칠 수 있다. 발표자 태도에 대한 적절한 피드백이 주어진다면 발표자의 발표 불안 증세는 많이 감소할 수 있으며 발표자는 발표 역량을 강화할 수 있다[4][5].

기존 연구에 따르면 발표자의 안정적인 발화와 시선 처리, 적절한 피드백이 발표자의 발표에 긍정적인 영향을 미친다는 보고가 있었다[6]. IT의 발전으로 타인이나 동료로부터의 피드백이 아닌 발표 동영상을 기술적으로 자동 분석해서 발표자의 발표 태도를 교정할 수 있게 되었다. 하지만 발표 연습을 도와주는 인간이 없어도 발표 습관을 스스로 개선할 수 있도록 도와주는 소프트웨어 프로그램은 면접 도우미 프로그램을 제외하면 거의 개발되고 있지 않다. 면접자의 말이나 행동에서 잘못된 부분을 확인하고 개선할 수 있도록 하는 면접 도우미 프로그램이 개발되어 있으나 활용 분야가 면접으로 국한되고 발표자의 태도를 확인하는 데에 제한이 있다.

본 논문에서는 발표자의 잘못된 발표 습관을 확인하

여 교정하는 기능을 제공하는 영상 기반 발표 태도 교정 프로그램을 개발한다. 본 연구는 사용자가 발표 영상을 업로드하면 분석하여 발표의 태도의 잘못된 부분을 지적한다. 예를 들어 영상에서 발표자의 안면을 분석하여 설정된 얼굴 탐지 알고리즘에 따라 정면 응시 여부를 체크하고 사용자에게 알려준다.

본 연구는 발표 태도 교정 보조 프로그램을 제안하기에 언어적·비언어적인 측면에서 청중을 고려한 발표란 무엇인지 분석한 연구[6]를 살펴보고, 발표 태도에서 비언어적 표현에 초점을 두었고, 연구 결과의 '어'와 같이 군말을 사용하는 경우 그리고 행동 분야에서 대부분 오래 보는 경우, 스크린을 너무 오래 보는 경우를 기준으로 진행하였다. 본 연구에서 개발한 프로그램을 사용하면 사용자는 본인이 원할 때 발표에 대한 피드백을 들을 수 있고, 최근 코로나 상황에서 타인과 만남이 꺼려지는 때에도 본 논문에서 개발한 프로그램을 이용한다면 타인의 도움이나 피드백 없이 객관적으로 발표 능력을 평가받고 발표 습관을 개선할 수 있다.

이를 위해 본 연구에서는 영상에서 발표자의 음성 인식을 위한 파이썬 기반의 API (Application Programming Interface), 발표자의 눈의 위치 등 안면 인식을 위한 API, 음성을 문자로 바꾸는 STT(Speech To Text) API를 통하여 발표자의 태도를 분석하고 발표자가 군말(redundant words)을 사용하고 있는지 발표자가 발표문이나 발표 스크린만을 응시하고 있지 않은지를 군말에 대한 필터(speech filter)를 이용하여 분석하고 판단해 주는 웹 기반의 프로그램을 개발하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 인공지능 기반의 영상과 음성 인식에 관한 기존 연구를 분석하고, 모의 면접 및 역량 검사 게임 앱인 셀프뷰(SelfView)[7]의 특성과 본 연구와의 차별점을 기술한다. 또한, 본 연구의 방향성을 기술한다. 3장에서는 본 연구에서 개발하고자 하는 프로그램의 설계 내용, 4장에서는 제안한 프로그램의 구현과 성능 분석 결과를 기술한다. 마지막으로 5장에서 본 연구의 제한점과 결론을 기술한다.

II. 연구 배경

1. 셀프뷰(SelfView)

셀프뷰는 실제적인 면접에 대비하여 스스로 자신의 면접을 연습하고 잘못된 부분을 교정할 수 있도록 도와주는 앱이다[7]. 인공지능을 이용하여 웹캠으로 얻은 면접자의 영상을 분석하고, 목소리 톤(tone)에 대한 분석, 어휘 분석, 표정 분석, 응답시간과 속도 등을 제시한다[7]. 셀프뷰는 본 연구에서 제안하는 프로그램과 유사하게 듣는 이의 관점에서 카메라 구도를 설정하고 녹화를 진행한다. 하지만 면접자와 카메라의 거리가 가까운 상황에서 면접자의 얼굴의 표정만을 가지고 분석하는 반면[8], 본 연구에서는 카메라와의 거리의 제약을 받지 않고 발표자의 태도 중 청중 응시 여부에 중점을 두어 분석한다.

셀프뷰는 각각의 면접 문항에 대해서 음성 데이터를 수집하며 해당 문항에 한하여 데이터를 분석한다. 본 연구는 발표의 상황에서 발표자의 음성을 분석하는데 내용의 특정한 항목에 제한되지 않고 행동 분석 영역도 상대적으로 넓은 거리에서 발표자의 태도를 분석하고 할 수 있도록 개선한다.

2. 얼굴 특징 분류기 고찰

인공지능 영역에서 얼굴 인식에 관한 연구는 많다[8-10]. 조병모 등의 연구에서는 카메라가 위치한 환경상의 최소의 샘플 이미지를 가지고, 그 환경에서 입력되는 영상의 인식 성공률을 높일 수 있는 적응형 얼굴 인식 방법을 마련하였다[9]. 또한, 이우리 등의 연구에서는 얼굴 표정에 관한 이미지 제작을 위해 인간의 6가지 감성 이미지를 제작하기도 하였다[10]. 이 절에서는 본 연구와 관련된 얼굴 특징 분류기를 중심으로 기존 문헌을 고찰한다.

2.1 HOG, SVM 분류기

HOG(Histogram of Oriented Gradient)는 영상에서 객체 검출에 사용하기 위해 개발되었다. HOG를 통해 객체에 대한 특징 정보를 추출하고, SVM(Support Vector Machine)으로 앞서 추출한 특징 정보를 통해 특정 객체 여부를 구분한다. 이 방법으로 얼

굴의 특징을 찾아 영상 내에서 얼굴인지를 파악한다. HOG는 대상 영역을 지정하고 일정 크기의 셀로 분할하고, 각 셀마다 에지(edge) 픽셀들의 방향에 대한 히스토그램을 구한 뒤 각 히스토그램의 이진 값을 선형으로 나타낸 벡터이다[11-13]. HOG는 영상의 밝기 변화에 덜 민감하고 내부 패턴이 복잡하지 않은 대상을 식별하는데 유리하다. 또한, 주로 윤곽선으로 물체를 식별하는 데 사용하기 적합하다.

2.2 LBP 분류기

LBP(Local Binary Pattern)은 영상에서 땅, 나무, 잔디 등 각 텍스처(texture)를 분류를 위해 개발되었다. 영상의 모든 픽셀에 대하여 3x3 영역으로 지정하고 상대적 밝기 변화를 밝으면 1, 어두우면 0으로 변환하여 인덱스 값으로 만든다. 이들에 대해 히스토그램을 구한 후, 이를 이용해 텍스처 모델로 활용하는 방식이다[14][15]. LBP를 이용해 영상 속의 얼굴 검출에도 활용할 수 있다. LBP는 영상의 밝기 변화에 강하며 어느 정도 내부 패턴이 존재하는 물체를 검출하기에 적절하다.

2.3 Haar-based cascade 분류기

Haar-based cascade 분류기는 영상 이미지에서 객체 검출(recognition)을 하는 머신러닝 알고리즘이다[16][17]. 사람 얼굴이 포함된 긍정적 이미지(positive image)와 얼굴이 포함되지 않은 부정적 이미지(negative image)를 훈련 시켜 샘플로부터 서로 다른 특징점들을 추출하여 객체를 분류한다[16][17]. 이를 통하면 [그림 1]과 같이 눈과 코, 전면 얼굴, 옆면 얼굴 등을 검출할 수 있다.

Haar-based cascade 분류기는 물체의 영역 간의 밝기 차를 이용하기 때문에 눈썹, 입술 등 특징적인 밝기를 가진 얼굴 검출에 적합하고 영역 내부에서 입술이나 눈썹 등의 형태 변화나 약간의 위치 변화에도 검출에 부정적 영향을 주지 않는다.

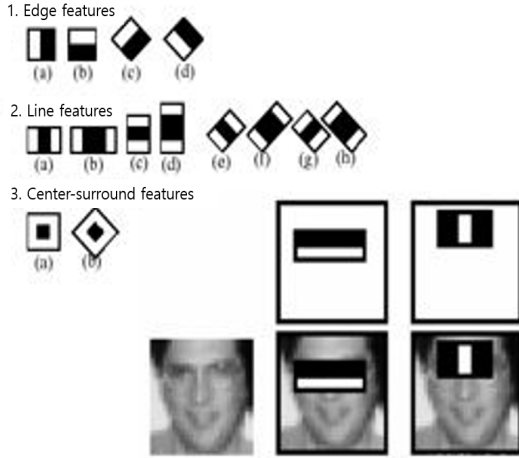


그림 1. Cascade classifier[16][17]

2.4 본 연구의 방향성

가버 필터(Garbor Filter)는 영상에서 객체의 외곽선 등 특징을 추출(extraction) 할 때 많이 활용되는 필터이다. 가버 필터는 영상의 특징점을 추출할 수 있어서 널리 사용된다. 김삼훈 등의 연구에서는 먼저 다운 샘플링된 입력 얼굴 이미지에서 초기 눈 좌표에서의 가버 특징 벡터와 해당 스케일의 눈 모델 번치와의 가버젯 유사도를 이용하여 눈 좌표를 추정한다[18]. 하지만 가버 필터가 객체 검출 시 넓은 배경이 포함된다는 문제점이 있다고 보고되고 있다[19].

한편, Ramos 등의 연구에서 얼굴 인식(recognition)은 Haar-Cascade 분류기를 사용하였고, 연구 감정을 파악하기 위한 얼굴 특징점 추출에서 가버 필터를 사용하여 성능을 높였다[20]. 본 연구에서는 머신러닝 기반 객체 검출 알고리즘인 Haar Cascade[16][17]를 사용하여 영상에서의 얼굴을 탐지하였다. 탐지된 얼굴에서 특징적인 부분(눈, 코, 입, 외곽선)에 좌표점을 설정하고, 점을 이어 영역을 구성하여 면적 비율 연산을 통해 얼굴의 정면 응시를 탐지하도록 한다.

3. 음성 인식과 합성곱 신경망을 이용한 단어 추출

최근 딥러닝 알고리즘은 우수한 성능을 드러내기 때문에 영상 인식과 자연어 처리 영역에서 많이 사용된다

[21][22]. 특히, 합성곱 신경망(Convolutional Neural Network, CNN)[23]은 여러 영역에서 활용되는 딥러닝 알고리즘이다. CNN은 주로 영상 인식에서 많이 활용되어 왔으나 언어 처리 영역에서도 활용되고 있다 [24][25]. 특히, 음성 특징점 추출을 위해 MFCC(Mel-Frequency Cepstral Coefficient)를 사용하고, 탐지 및 분류를 위해 CNN을 사용한 연구가 있다[26]. 신경식 등의 논문에서는 CNN을 활용하여 연산량을 줄여 성능의 향상을 가져왔다[26]. 본 연구에서도 MFCC를 이용하여 목소리를 분석하고, CNN을 이용하여 군말을 인식한다. 본 연구에서도 이와 같은 연구 결과를 바탕으로 파이썬의 라이브러리인 librosa[27]를 기반으로 MFCC와 CNN을 이용하여 군말을 인식하였다.

III. 연구 설계

본 연구에서 진행한 내용은 다음 [표 1]과 같다. 군말 사용 여부 확인, 안면 인식, 발표 내용 텍스트화, 웹 프로그램 인터페이스 구현을 통해 발표 태도를 지능적으로 교정할 수 있도록 한다.

표 1. 개발 항목에 관한 설명

개발 항목	항목에 관한 설명
군말 여부 확인	합성곱 신경망을 사용하여 군말 데이터 셋에서 추출된 정보를 학습시키고 군말을 탐지하는 인공지능 모델을 만든다. 인공지능 모델을 이용하여 영상 속 발표자의 목소리를 분석하고 발표자의 발화에서 군말(어, 음, 그)을 사용하였는지 확인한다.
안면 인식	안면 인식 API를 이용하여 발표자 얼굴의 눈, 코, 입, 외곽선을 이용한 기준을 설정한다. 영상 속의 발표자가 청중과 시선을 마주치지 않고 발표문 혹은 발표 화면을 설정된 시간 이상 보면서 발표를 진행했는지 얼굴 탐지 알고리즘을 설계 및 구현하여 탐지한다.
발표 내용 텍스트화	Google STT API와 군말 사용 여부 확인 인공지능 모델을 이용하여 영상 속 발표자의 발표 내용을 텍스트로 변환시킨다. 텍스트와 발표문을 비교하여 발표가 발표자의 의도대로 진행되었는지 판단할 수 있도록 도와준다.
웹 프로그램 인터페이스 구현	HTML5, CSS, JavaScript를 이용하여 동영상을 업로드 받고 분석하여 웹사이트 내에서 재생할 수 있도록 한다. 사이트 내 동영상의 타임라인을 통해 사용자에게 앞서 찾아내었던 군말 확인 결과와 안면 인식 판단 결과를 표시한다. 그리고 텍스트화한 발표 내용을 타임라인 하단에 출력한다.

본 연구에서 개발한 프로그램에서 군말 여부 확인 과정은 다음 [그림 2]와 같다. 군말 여부 확인 과정은 음성 인식 과정, 군말 사용 확인 과정, 결과물 도출로 나

누어 진행했다. 먼저 음성 인식 과정에서는 입력된 영상에서 음성을 추출할 수 있도록 하였다. 군말 여부 확인 과정에서는 군말 데이터셋을 미리 학습시킨 군말 분류 인공지능 모델을 이용하여 음성에서 사용한 군말을 분류하도록 하였다. 하지만 앞서 음성에서 분류한 군말에는 문장에서 의도적으로 사용한 군말이 존재한다. 군말 사용 판단 인공지능 모델을 이용해 분류한 음성이 의도적으로 사용한 것인지 불필요하게 사용한 군말인지 판단하도록 하였다. 결과물 도출 과정에서는 추출한 음성에 대해 전사화를 진행하였고 불필요하게 사용한 군말을 다른 색깔로 칠하여 체크할 수 있도록 하였다. 해당 과정의 결과물은 군말이 탐지된 시점을 타임라인에 표시한 마커와 그에 대응되어 문제점과 해결을 위한 피드백을 제시한다. 또한, 각각의 군말 탐지 횟수와 군말이 하이라이트 처리된 음성 전사 파일을 제공한다.

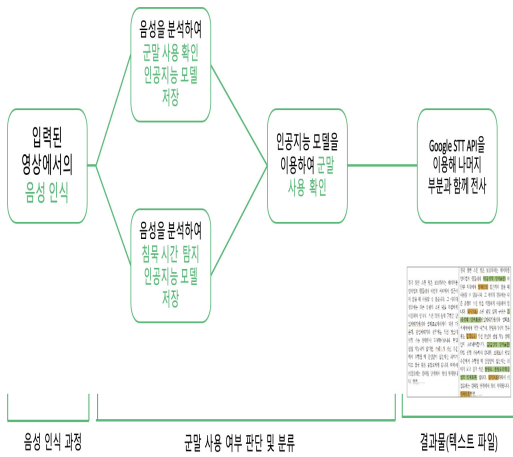


그림 2. 군말 확인 및 영상의 음성 전사화

또한, 얼굴의 정면 응시 여부 탐지 과정은 다음 [그림 3]과 같다. 얼굴의 정면 응시 여부 탐지는 크게 얼굴 인식 과정, 정면 여부 탐지 과정, 결과물 도출로 나누어 진행했다. 먼저 얼굴 인식에서는 입력된 영상에서 얼굴을 인식하고 해당 얼굴을 토대로 특징점을 찍어 이후 정면 여부를 판단할 수 있도록 하였다. 그리고 정면 여부 탐지 과정에서는 특징점을 이용하여 옆면과 아랫면의 탐지 기준을 설정하고 제한조건을 넘어가면 탐지하도록 설계하였다. 탐지된 결과는 시간, 내용, 횟수를 포

함하도록 하였다. 해당 과정의 결과물은 문제가 되는 얼굴 행동이 인식된 시점을 타임라인에 표시한 마커와 그에 대응되어 문제점, 해결 방안, 얼굴의 옆면과 아랫면 탐지 횟수를 제공한다.

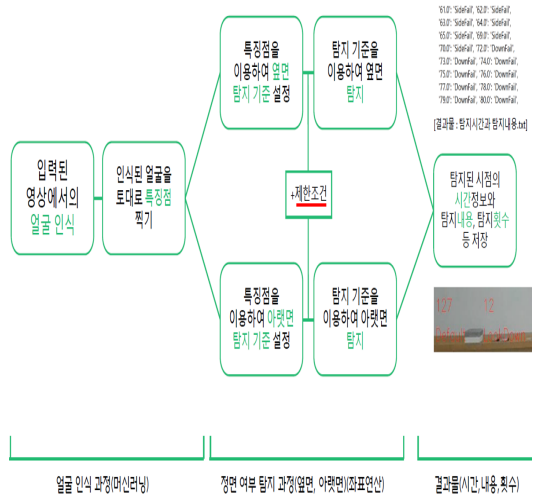


그림 3. 얼굴의 정면 응시 여부 탐지

IV. 구현과 성능 분석

1. 구현 환경

본 연구에서 개발한 앱의 구현 환경은 [표 2]와 같다. 본 연구에서 개발한 프로그램은 웹사이트 형식으로 영상을 업로드하면 해당 서비스를 제공한다. HTML, CSS, JavaScript와 같은 웹 언어와 python 프로그래밍 언어를 이용하여 앱을 구현하였다. 웹 기반으로 프로그램을 구현하면 다른 응용 프로그램과 달리 추가로 파일을 설치하거나 다운로드하지 않고 웹상에서 직접 서비스를 이용할 수 있는 이점이 있다.

표 2. 구현 환경

프로그램 구현	사용된 도구
프로그램 개발에 사용할 언어	웹 언어(HTML, CSS, Javascript), 파이썬(Python)
운영체제(OS)	Windows 10
통합 개발 환경(IDE)	Visual Studio Code, Colab
음성 인식	Google Cloud Speech-to-Text, Librosa
영상 인식	numpy, Opencv

Google Cloud STT를 통해 발표 영상의 음성 데이터를 이용하여 텍스트 형태의 스크립트로 변환하였다. Librosa 라이브러리 같은 경우 영상에서 음성 데이터를 추출하였다. numpy와 Opencv 라이브러리를 통해 haar cascade를 이용하여 눈, 코, 입을 인식하고 인식된 특징점을 연산하여 발표자가 청중을 제대로 응시하고 있는지 판단하였다.

2. 음성 및 영상 구현 사항 및 분석

군말 여부 확인 기능 구현 과정은 다음과 같다. 영상에서 Librosa 라이브러리를 활용하여 음성의 특징을 추출한 후 군말을 사용한 부분을 찾아내었다. 군말을 사용하지 않은 부분은 바로 Google의 STT API를 활용해서 전사하였다. 본 연구에서 판단할 수 있는 군말은 음, 그, 어이다. 군말을 사용한 부분은 음, 그, 어를 학습시킨 인공지능 모델을 이용하여 따로 전사시키고, 이후 Google의 STT API를 활용한 전사 데이터와 군말 판단 인공지능 모델을 통해 전사한 데이터를 합치도록 설계하였다.

결과적으로 두 가지 인공지능 모델을 만들어야 하는데 첫 번째는 어떤 군말을 사용했는지 분류해 주는 인공지능 모델이고, 두 번째는 문장에서 군말을 판단하는 인공지능 모델이다. 첫 번째 인공지능 모델은 구현하였지만 두 번째 인공지능 모델은 구현하지 못하였다. 문장에서 군말 사용을 정확도 있게 잡아내기 위해서는 그 단어의 음성적 데이터만으로 판단하는 것이 아니라 군말에 대한 기준을 잡아야 하는데, 여러 가지 기준을 설정하여 시도해 봤지만, 정확도 있게 잡아낼 수가 없었다. 예를 들어 '그림'이라는 단어에서의 '그'와 군말의 '그'를 구분하는데 오류가 발생하였다.

다음은 위에서 말한 첫 번째 인공지능 모델을 구현하면서 사용한 음성 인식 알고리즘이다. 음성에 들어 있는 정보는 음성 신호 자체에서 쉽사리 얻어낼 수 없어서, 수학적인 신호 처리를 거쳐야만 추출할 수 있다. 대표적으로 음성을 주파수라는 또 다른 축으로 관측하는 방법이 있는데 본 인공지능 모델은 이 주파수를 이용하였다. 푸리에 변환 함수를 사용해서 특정 시간 길이의 음성 조각이 각각의 주파수 성분들을 얼마만큼 가졌는지를 의미하는 스펙트럼을 얻을 수 있었고, 이렇게 얻

은 여러 개의 스펙트럼을 시간 축에 나열하면 시간 변화에 따른 스펙트럼의 변화인 스펙토그램을 얻을 수 있었다. 얻은 스펙토그램을 이용하여 일련의 과정을 거쳐 MFCC를 추출하였다. 이상의 특징 추출에는 Librosa 라이브러리를 사용하였다. MFCC는 음향과 음성 인식에서 자주 사용되는 특징 벡터 추출 알고리즘이다. 여러 음성 특징 벡터 추출 알고리즘 중 연산량 대비 높은 성능의 특징 벡터를 추출할 수 있다는 장점을 가지고 있다[28]. 이 특징 벡터를 이용해 컨볼루션 계층과 풀링 계층을 쌓아가며 합성곱 신경망을 만들어 인공지능 모델을 만들었다[29].

[그림 4]는 군말 분류 인공지능 모델의 정확도이다. Keras의 model.fit()함수를 통해 훈련 집합(train set)의 20%를 분리하여 검증 데이터로 사용하여 학습시켰고, 이에 대한 정확도를 얻었다. 이후 파이썬의 시각화 패키지인 matplotlib의 pyplot 모듈을 통해 에포크의 증가에 따른 정확도를 그래프로 표현하였다. 본 모델은 약 75%의 정확도를 가지고 있다.

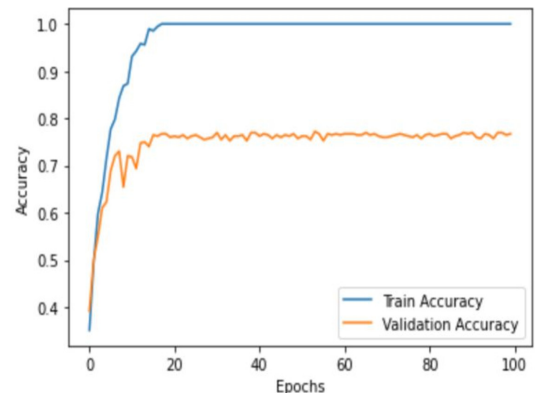


그림 4. 군말 분류 인공지능 모델

선행 연구에 제시되었던 발표자의 발표 태도에 대한 잘못된 유형 중 영상 인식으로 구현 가능한 부분은 프로젝트 스크린을 보고 발표하거나 발표문만을 보고 발표하는 상황과 같이 청중을 응시하지 않는 태도를 탐지하는 것이다. 그래서 기능을 크게 두 가지로 나누어 프로젝트 스크린을 보는 상황에서는 발표자의 고개가 옆을 향해 있는지, 발표문을 보는 상황에서는 고개가 아래를 향해 있는지를 탐지하고 일정 시간 이상 각각의

태도를 보일 시 그 횟수를 카운트하도록 구성하였다.

먼저 얼굴의 특징을 학습하는 단계는 만개가 넘는 사진 데이터를 가지고 머신러닝을 통해 반복적으로 명암을 분석하면서 haar cascade 기법에 따라 얼굴 특징을 찾아내는 기준을 만들도록 한다. 이때의 기준은 XML 형태로 저장된다. XML 형태의 기준은 얼굴 외곽선과 얼굴 내 요소인 눈, 코, 입, 눈썹 등으로 나누어 저장했다. 이 단계에서 만들어진 기준을 가지고 다음 단계에서 입력된 사진과 반복적으로 비교하며 얼굴의 위치가 어디인지 찾아낸다. 해당 얼굴의 위치에 눈, 코, 입, 눈썹, 얼굴 외곽선 등 특징점을 찍는다. 얼굴의 위치를 특정하지 않고 사진에 바로 얼굴의 특징점을 찍어도 되지만 정확성을 높이기 위해 얼굴의 위치를 특정하여 [그림 5]와 같이 특징점을 찍었다. [그림 5]는 인공지능이 생성한 얼굴 이미지이다[30].

이렇게 찍어놓은 특징점을 활용한 연산을 하여 발표자의 얼굴이 얼마나 돌아갔는지를 판단할 수 있다. 앞서 얼굴 인식에 머신러닝을 사용하였지만, 발표자의 옆면과 아래면 응시 여부를 인공지능을 사용하지 않고 특징점 연산만으로 처리하였다.

본 연구는 옆면과 아래면 응시에 더불어 추가적인 기능을 제공하지 않고 옆면과 아래면 인식을 할 수 있으면 충분하며, 인공지능 연산보다 특징점 좌표연산이 연산량이 적어 프로그램의 수행 시간에 차이를 발생시킨다. 그래서 좌표점을 이용한 연산으로 옆면과 아래면을 탐지하는 알고리즘을 설계하고 구현하였다.

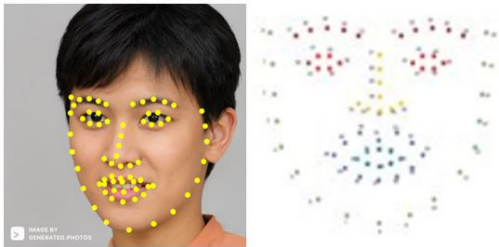


그림 5. Haar feature[30]

옆면 탐지에서는 각 특징점을 모아 하나의 영역으로 구성하고 사전 순으로 알파벳을 붙여 면적을 계산하였다. [그림 6]의 왼쪽은 얼굴이 정면을 향해 있을 때를

초록색으로 표시한 것이고 오른쪽은 얼굴이 옆을 향해 있을 때를 빨간색으로 표시한 것이다. 얼굴의 회전에 따라 벡터의 외적을 이용하여 각 영역의 면적을 구하고 비율을 비교하였다. 영역 a와 b를 예로 들면 $a = \text{point}(0, 2, 36)$, $b = \text{point}(14, 16, 45)$ 로 구성하고,

$$\left(|(x2-x1)(y3-y1)-(x3-x1)(y2-y1)| \right) / 2 \quad (1)$$

식(1)을 통해 영역을 구하였다. 식에서 $x1, x2, x3$ 는 포인트 a의 좌표점, $y1, y2, y3$ 는 포인트 b의 좌표점을 일반화한 것이다. 그리고 각 영역을

$$\text{ratio}((a-b)/(a+b)) \quad (2)$$

식(2)을 통해 비율을 계산하여 양수이면 발표자 기준으로 왼쪽을 보았을 때, 음수이면 오른쪽을 보았을 때로 탐지하도록 설계하였다. 얼굴을 살짝만 회전하여도 옆면으로 볼 수 있으므로, 가장 옆면일 때의 점 간의 상대적 위치를 제한조건으로 두어 옆면의 기준을 설정하였다. 그리고 제한조건에 따라 옆면 인식이 되면 일정 시간 동안 해당 상태를 유지할 시 카운트하도록 하였다. 여기에서는 프로젝터 스크린을 이용하는 상황을 가정하여 5초를 기준으로 일정 시간을 결정하였다.

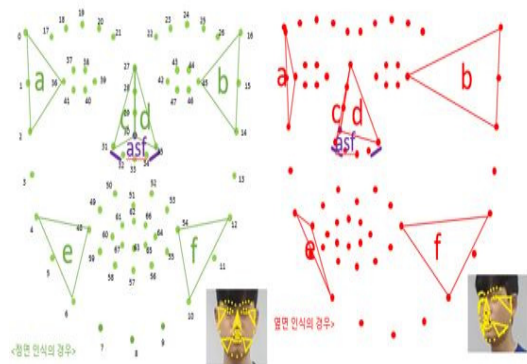


그림 6. 옆 응시 탐지를 위한 영역 구성

얼굴의 아래 응시도 마찬가지로 좌표를 가지고 연산하였고 상대 거리를 이용하여 탐지하였다. [그림 7]의

왼쪽은 얼굴이 정면을 향해 있을 때를 초록색으로 표시한 것이고 오른쪽은 얼굴이 아래를 향해 있을 때를 빨간색으로 표시한 것이다. 얼굴의 고갯짓에 가장 민감하게 반응하는 점을 기준으로 잡고 사전 순으로 알파벳을 붙여 연결하여 거리를 계산하였고 정면을 응시할 때와 아래를 응시할 때 나타나는 특징을 이용하여 탐지하였다. 이때에도 어느 정도 고개를 아래로 내렸을 때와 아래 응시로 탐지할 것인지 구별하기 위해 제한조건을 설정하여 탐지하였다. 아래 응시 탐지에서 제한조건을 충족시키는 식은 식(3)과 같다.

$$(y(\text{point}(3)) + y(\text{point}(13))) / 2 \quad (3)$$

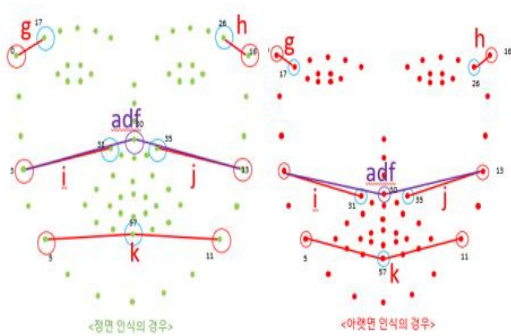


그림 7. 아래 응시 탐지를 위한 상대 거리 구성

그리고 제한조건에 따라 아랫면 인식이 되면 일정 시간 동안 해당 상태를 유지할 시 카운트하도록 하였다. 여기서는 발표자 앞에 발표문이나 컴퓨터 화면이 있는 발표 상황을 가정하여 5초를 기준으로 일정 시간의 정도를 결정하였다.

영상에서 얼굴 탐지 결과에 대해 정량적으로 측정하여 결과를 도출하였다. 테스트 영상은 5분, 10분, 15분 발표의 상황을 가정하고 각 시간에 맞게 영상을 준비하였다. 영상의 내용은 앞을 바라보며 발표하는 모습, 스크린을 바라보고 발표하는 모습, 발표문을 바라보고 발표하는 모습을 세 가지로 구성하였고 두세 번째 경우에 대하여 일정 시간보다 적게 바라보는 경우, 일정 시간 이상 바라보는 경우를 포함하도록 하였다. 각 테스트 영상에서 모범 데이터를 기준으로 구현한 얼굴 탐지 기능이 얼마나 정확도를 가지는지 측정하였다. 그리고 구

현한 좌표연산 알고리즘과의 연산속도, 정확도 비교를 위해 Teachable Machine의 영상 인식 인공지능 모델을 이용하였다[31]. Teachable Machine은 웹 기반 머신러닝 도구로 이미지를 학습시켜 쉽고 빠르게 모델을 생성할 수 있게 하는 도구이다. 영상 인식 모델에 약 4,000개가량의 얼굴 사진(옆 응시, 앞 응시, 아래 응시)을 epoch 50, batch size 16에 따라 학습시켜 얼굴 탐지 인공지능을 만들었다. 얼굴 탐지 인공지능 역시 좌표연산 알고리즘과 같은 테스트 영상에 대해 탐지하도록 하였다.

[그림 8]은 영상 길이에 따라 모범 데이터에 대하여 앞서 구현한 좌표연산 알고리즘과 얼굴 탐지 인공지능이 얼마나 정확성을 가지는지 보여준다. 5분 길이의 영상에서는 좌표연산 알고리즘이 35%, 얼굴 탐지 인공지능이 34%의 정확도를 가졌고, 10분 길이에서는 각각 22%와 19%, 15분 길이에서는 각각 20%와 5%를 보였다. 전체적으로 영상 길이가 길어질수록 정확도가 낮아지는 양상을 보인다. 프레임당 탐지 속도를 비교해보면 평균적으로 좌표연산 알고리즘은 0.005초, 얼굴 탐지 인공지능은 0.09초로 측정이 되었다.

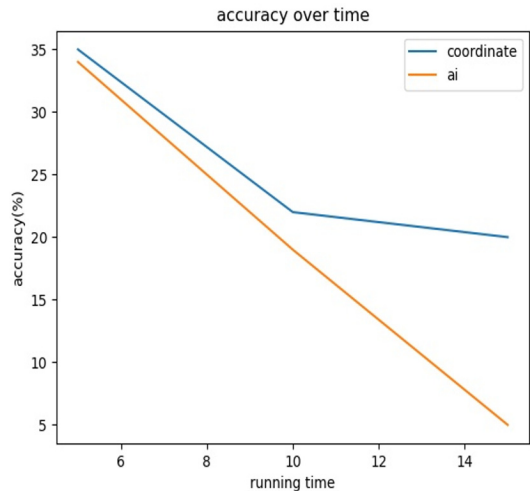


그림 8. 영상 길이에 따른 정확도 그래프

[그림 9]는 영상 길이에 따라 좌표연산 알고리즘과 얼굴 탐지 인공지능이 얼마나 빠른 탐지 속도를 가지는지 보여준다. 5분 길이의 영상에서는 좌표연산 알고리

즘이 31.49초, 얼굴 탐지 인공지능이 727.41초의 탐지 속도를 보였고, 10분 길이에서는 각각 63.13초와 1498.79초, 15분 길이에서는 99.79초와 1880.69초를 보였다. 전체적으로 영상 길이가 길어질수록 탐지 시간이 누적되어 전체적인 탐지 속도가 커지는 양상을 보인다. 좌표연산 알고리즘은 얼굴 탐지 인공지능의 탐지와 달리 얼굴 위치에 적은 좌표점들을 가지고 영역을 구성하여 연산한 것이기 때문에 연산량이 적어 탐지 속도에서 우위를 보인다.

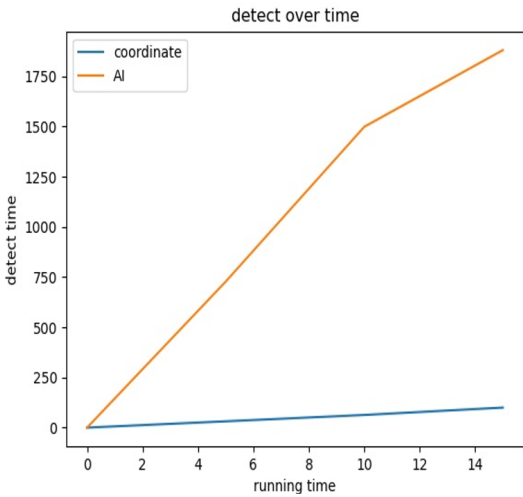


그림 9. 영상 길이에 따른 누적 탐지 속도 그래프

프로그램의 GUI는 다음 [그림 10]과 [그림 11]과 같이 구성하였다. GUI를 구성할 때 사용자가 영상 전체를 훑어보지 않고도 어디가 문제점이 있었고 어떤 문제였는지 편리하고도 쉽게 알아볼 수 있도록 설계하였다. 그림 10은 사용자가 자신이 발표한 영상을 업로드할 수 있는 페이지이다. 사용자는 발표 영상을 '파일 선택하기' 버튼을 누르거나 버튼 위의 점선 박스 안에 끌어당겨 업로드할 수 있다. 발표 영상을 업로드하면 '분석하기' 버튼을 눌러 영상 분석 페이지로 이동할 수 있다. [그림 11]은 발표 영상을 분석하여 사용자에게 보여주는 페이지이다. [그림 11]의 왼쪽 위는 발표 영상을 재생할 수 있는 영상 스크린이 위치에 있고 영상 스크린 아래에는 영상에 대한 타임라인이 구성되어 있다. 타임라인은 총 3개로 구성되어 있으며 중간 타임라인

(두 번째 타임라인)은 영상의 기본 재생 타임라인이다. 첫 번째 타임라인은 군말 탐지 결과에 대해 빨간색 마커를 통해 사용자에게 해당하는 군말의 위치를 표시해 준다. 세 번째 타임라인은 발표 영상의 얼굴 분석 결과에 대해 파란색 마커를 통해 사용자에게 문제가 있는 탐지 위치를 표시해 준다. 오른쪽 리스트는 첫 번째와 세 번째 타임라인에 표시된 마커에 대응하여 탐지된 문제점에 대한 설명을 보여준다. 아래의 스크립트는 발표 영상에서 음성을 전사화한 것이고 탐지한 군말은 탐지된 위치 사이사이에 빨간색 글씨와 하이라이트(어:노랑, 음:초록, 그:파랑) 처리하여 보여준다.

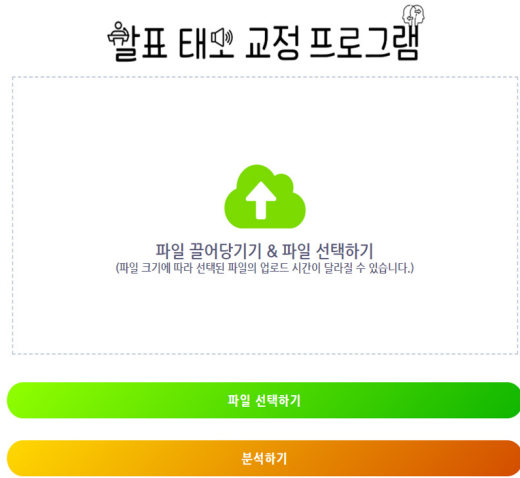
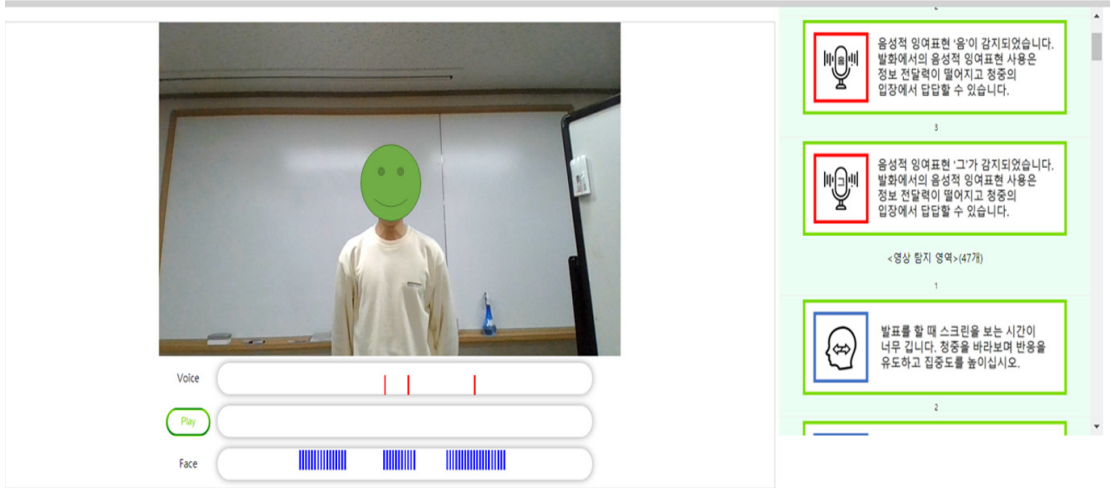


그림 10. 영상 업로드 페이지

V. 결론

결론적으로 영상에서의 얼굴 분석 기능에 대해서는 모범 탐지 데이터와 비교하여 좌표연산 알고리즘이 부족한 정확도를 보였으나 같이 비교한 얼굴 탐지 인공지능에 비해 높은 정확도를 보였다. 또한, 순수 탐지 시간을 비교하였을 때, 좌표연산 알고리즘은 얼굴 탐지 인공지능과 비교해 연산 속도 면에서 우세하였다. 아직 인공지능이 학습한 데이터양이 약 4,000개로 부족하고 학습 데이터의 질이 낮아 현재 상황으로써 좌표연산 알고리즘이 우세한 양상을 보여주지만 이후 얼굴 탐지 인

출표 태도 교정 프로그램



제주대학 홍길동 교수는 헌법 13조라는 관점에서 문제를 제기하고 있습니다 일본의 헌법 13조에 따르면 모든 국민은 개인으로서 존중받아야 권리가 있다고 쓰여 있는데요 디지털 학습을 한 ai로 개인의 실체가 개인의 그런 정보에 점점 더 접근해야하는 프로파일링 기술이 점점 더 발전하고 있습니다 이는 인재채용을 할 때나 신용등급을 **안(주임새)** 파악하는 등 금융 서비스에서도 활용이 가능하다 가능하면서 **음(주임새)** 응용 영역이 대단히 넓습니다 그러나 이런 식으로 개인을 분류하기 시작하면 시작하면 점점 개인의 이런 정보를 기반으로 개인을 차별하는 시대가 다가올 수도 있다고 홍길동 교수는 **그(주임새)** 우리가 시를 통해서 얻을 수 있는 것은 디지털화를 통해서 예측을 하고

그림 11. 영상 분석 페이지

공지능이 더 많은 데이터를 학습한다면 정확도에서는 상대적으로 떨어지겠으나 여전히 연산 속도에서 우세할 것이다.

본 논문은 영상에서의 얼굴 회전을 탐지하기 위해 알고리즘을 설계하는 과정을 통해 인공지능과 비교하여 성능이 비슷하거나 높으면서도 탐지 속도를 높이려 노력한 점에 의의가 있다. 얼굴 검출에서는 머신러닝으로 학습된 Haar-based cascade 분류기를 이용하였지만, 이후 검출 영역에서 특징점을 통한 얼굴의 회전 탐지는 벡터의 외적을 통해 비교 영역을 구성하고 상, 하, 좌, 우의 회전 정도를 계산하여 4,000개가량의 데이터가 학습된 인공지능을 사용했을 때보다 신경망에 따른 연산량을 획기적으로 줄이면서도 성능을 높이는 결과를 보였다.

현재 연구에서는 음성 탐지 영역에서 설정해 놓은 어, 음, 그를 탐지할 수 있지만 탐지한 어, 음, 그가 군말인지 '그림'의 '그'와 같이 필요한 단어인지 명확히 분별하지 못한다는 점에서 한계가 있다. 그리고 영상 탐지 영역에서는 얼굴의 회전 정도를 고려할 때 거리에 따른 얼굴 탐지의 한계로 얼굴 영역의 눈동자도 같이

고려하지 않고 특징점을 통해 지정한 각 영역의 면적 비율만을 고려한 점에서 한계가 있다. 얼굴이 회전된 상태에서 일정 시간을 결정할 때 다양한 발표 상황을 고려하지 않고 단편적 기준만으로 결정한 점에서 한계가 있다. 정확도가 모범 데이터와 비교해 떨어지는 점에서 한계가 있다.

앞선 한계점을 보완하고 기능을 확장한다면 음성 탐지 영역에서 더 정확한 군말 분별을 통해 발표자에게 자주 사용하는 군말 등 통계를 내어 제시할 수 있고, 좀 더 넓은 범위에서의 말실수를 학습시켜 발표 상황뿐만 아니라 발화에 익숙지 못하거나 어려움을 겪는 아동의 언어치료에도 활용할 수 있을 것이다. 그리고 영상 탐지 영역에서는 단순히 옆을 보거나 아래를 일정 시간 이상 본다는 것만으로 상황을 단편적으로 해석하여 탐지하지 않고 좀 더 다양한 상황을 조사해 조건을 추가로 구성하여 발표 태도 탐지를 할 수 있을 것이다. 예를 들어 발표 중 스크린에 제시된 그림을 가리키면서 설명하는 경우 옆을 일정 시간 이상 볼 수 있으나 이는 발표 태도에 있어서 문제가 된다고 볼 수 없다. 이 상황을 고려하여 유연하게 탐지함으로써 발표자에게 정확한 탐

지 결과를 제공할 수 있다. 그리고 본 논문에 적용한 얼굴 탐지 알고리즘처럼 얼굴 특징점을 이용해 영역을 구성하고 이들 영역을 이용해 비율을 구한다면 연산 방식을 다른 분야에도 적용 가능하다. 예를 들어 얼굴 특징점을 통한 영역을 약간 달리 구성하여 얼굴 내 요소끼리 엮어 놓고 각 영역의 비율을 연산한다면 일반적인 표정 패턴에 따른 눈, 입술, 눈썹의 위치나 모양 비율에 따라 표정 탐지에도 적용할 수 있는 등 얼굴 내 영역을 통한 다양한 사례에 활용할 수 있다.

본 논문에서 발표자의 잘못된 발표 습관을 확인하여 교정하는 기능을 제공하는 웹 기반의 영상 기반 발표 태도 교정 보조 앱을 개발하였다. 발표 태도에 대한 피드백은 발표자가 자신의 습관을 자각하게 하고, 발표 태도의 변화를 주는 계기가 된다. 발표자는 본 논문의 발표 태도 교정 앱으로 제공되는 피드백을 통해 잘못된 습관을 고쳐 발표에 자신감을 얻을 수 있다. 그리고 발표자 태도에 대한 적절한 피드백이 주어진다면 발표자의 발표 불안 증세는 많이 감소할 수 있으며[5][32] 발표자는 발표 역량을 강화할 수 있다.

참 고 문 헌

[1] <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=89671&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=020402&opType=N>

[2] 송인섭, “학습자 중심의 21세기 패러다임: 방법과 전망,” *교육심리연구*, 제22권, 제4호, pp.881-896, 2008.

[3] 윤여경, “효과적인 의사소통을 위한 스피치 교육 앱 디자인 연구,” *상품문화디자인학연구*, 제66권, pp.43-52, 2020. <https://doi.org/10.18555/kicpd2020.60.05>

[4] 도원영, “말하기에서의 동작언어에 대한 고찰,” *한국어학*, 제39권, pp.191-221, 2008.

[5] 변정민, “청중을 고려한 발표자의 언어적·비언어적 표현 연구,” *새국어교육*, 제81권, pp.133-158, 2009.

[6] 김예지, *대학생 발표 불안 연구*. 전남대학교 교육대학원, 석사학위논문, 2012.

[7] <https://play.google.com/store/apps/details?id=com.enhanceu.selfview&hl=ko&gl=US>

[8] 송재민, 이새분, 박아름, “이미지 인식 기술의 산업 적용 동향 연구,” *한국콘텐츠학회논문지*, 제20권, 제 7호, pp.86-96, 2020.
<https://doi.org/10.5392/JKCA.2020.20.07.086>

[9] 조병모, 김기한, 이필규, “필터 및 특징 선택 기반의 적응형 얼굴 인식 방법,” *한국콘텐츠학회논문지*, 제9권, 제6호, pp.1-8, 2009.

[10] 이우리, 황민철, “한국인 표준 얼굴 표정 이미지의 감성 인식 정확률,” *한국콘텐츠학회논문지*, 제14권, 제9호, pp.467-483, 2014.
<https://doi.org/10.5392/JKCA.2014.14.09.476>

[11] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Vol.1, IEEE, 2005.

[12] J. Chen, Z. Chen, Z. Chi, and H. Fu, “Facial Expression Recognition Based on Facial Components Detection and HOG Features,” *Proceedings for the 2014 Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields*, pp.884-888, 2014.

[13] 유원필, *실외환경에 강인한 도로기반 저가형 자율주행기술개발*, 한국전자통신연구원, 2015.

[14] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *Proceedings of the 2006 IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, Vol.28, No.12, pp.2037-2041, 2006.

[15] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, “Learning Multi-scale Block Local Binary Patterns for Face Recognition,” *International Conference on Biometrics*. Springer, Berlin, Heidelberg, 2007.

[16] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001. Vol.1, pp.I-I. 2001.

[17] https://docs.opencv.org/4.x/d2/d99/tutorial_j

s_face_detection.html

[18] 김상훈, 정선태, 정수환, 오두식, 김재민, 조성원, “다중 스케일 가버 특징 벡터 모델 기반 눈좌표 검출,” 한국콘텐츠학회논문지, 제7권, 제1호, pp.48-57, 2007.

[19] 김성영, “가버 필터에 기반한 관심 객체 검출,” 한국컴퓨터정보학회논문지, 제13권, 제2호, pp.87-94, 2008.

[20] A. L. A. Ramos and B. G. Dadiz, “A Facial Expression Emotion Detection using Gabor Filter and Principal Component Analysis to Identify Teaching Pedagogy,” 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp.1-6, 2018, doi: 10.1109/HNICEM.2018.8666274.

[21] 이형일, 김진명, 이재원, “딥러닝 객체인식을 통한 경로보정 자율 주행 로봇의 구현,” 한국콘텐츠학회논문지, 제19권, 제12호, pp.164-172, 2019.

[22] 이도연, 장병희, “딥러닝을 이용한 음악흐름 예측모델 개발 연구,” 한국콘텐츠학회논문지, 제20권, 제8호, pp.10-18, 2020.

[23] http://wiki.hash.kr/index.php/합성곱_신경망

[24] 모경현, 박재선, 장명준, 강필성, “단어와 자소 기반 합성곱 신경망을 이용한 문서 분류,” 대한산업공학회지, 제44권, 제3호, pp.180-188, 2018.

[25] 박현정, 송민채, 신경식, “CNN을 적용한 한국어 상품평 감성분석: 형태소 임베딩을 중심으로,” 지능정보연구, 제24권, 제2호, pp.59-83, 2018.

[26] 신경식, 유신우, 오혁준, “MFCC와 CNN을 이용한 저고도 초소형 무인기 탐지 및 분류에 대한 연구,” 한국정보통신학회 논문지, 제24권, 제3호, pp.364-370, 2020.

[27] <https://librosa.org/doc/latest/index.html>

[28] N. Dave, “Feature Extraction Methods LPC PLP and MFCC in Speech Recognition,” Proceeding of the International Journal for Advance Research in Engineering and Technology, Vol.1, pp.1-5, 2013.

[29] <https://tech.kakaoenterprise.com/66>

[30] <https://generated.photos/>

[31] <https://teachablemachine.withgoogle.com/>

[32] 이종익, *발표능력 향상 프로그램이 여중생의 발표불안 감소 및 자기효능감 향상에 영향을 미치는 효과*, 강원대학교 교육대학원, 석사학위논문, 2009.

저 자 소 개

권 기 현(Kihyeon Kwon)

준회원

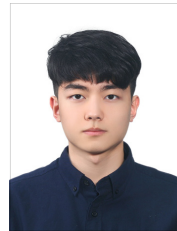


■ 2019년 3월 ~ 현재 : 제주대학교 컴퓨터교육과 학부생

<관심분야> : 정보교육, 에듀테크, 교육과정

안 수 호(Suho An)

준회원



■ 2018년 3월 ~ 2019년 12월 : 제주대학교 국어국문학과 학부생
 ■ 2020년 3월 ~ 현재 : 제주대학교 컴퓨터교육과 학부생

<관심분야> : 정보교육, 국어교육, 에듀테크

박 찬 정(Chan Jung Park)

종신회원



■ 1988년 2월 : 서강대학교 전자계산학과 (공학사)
 ■ 1990년 2월 : KAIST 전산학과(공학석사)
 ■ 1998년 2월 : 서강대학교 대학원 전자계산학과 (공학박사)
 ■ 1998년 2월 ~ 1999년 9월 : 한국

통신 멀티미디어연구소 전임연구원

■ 1999년 9월 ~ 현재 : 제주대학교 컴퓨터교육과 교수
 ■ 2013년 4월 ~ 2015년 2월 : 제주대학교 교육과학연구소 소장

<관심분야> : STEAM, 기술기반 교육, 에듀테크, 빅 데이터 처리