

온라인 화상 면접 연습 시스템

Online Video Interview Practice System

강유림, 박지은
(Yurim Kang and Jieun Park)

요약 : 최근 언택트 시대에 접어들면서 화상 면접 채용 제도를 실행하는 기업이 증가함에 따라 이를 대비할 필요성이 대두된다. 따라서 본 논문에서는 면접자의 비언어적 표현을 분석하여 이에 대한 피드백을 제공함으로써 온라인 화상 면접을 연습할 수 있는 시스템을 제시한다. OpenCV, Dlib, Matplotlib를 활용하여 Face Detection, Gaze Tracking, Facial Expression를 판단하고 이를 통해 면접자의 태도를 분석한다. 최종적으로 설계한 알고리즘을 하나의 UI로 제작하여 면접자에게 도움이 되는 시스템을 구현한다.

Abstract : As recruitment using online interview systems becomes popular in the recent untact era, there are growing needs to be fully prepared for the online, non-face-to-face job interview. In this paper, we propose a system that help interviewers practice online video interview by analyzing non-verbal expression and providing feedback. By using OpenCV, Dlib, and Matplotlib library, face detection, gaze tracking, and facial expression are performed. Finally, the developed algorithm is implemented in a UI that provides real-time feedback and analysis report to the interviewers.

Keywords: Online Interview, Face Detection, Facial Landmark, Gaze Tracking, Facial Expression

1. 서론

최근 COVID-19가 확산됨에 따라 채용 시장에서 비대면 화상 면접과 관련된 관심이 증가하면서, 그에 따라 화상 면접 채용 제도를 실행하는 기업이 증가하고 있다. 비대면 화상 면접이란 화상 통화 어플리케이션을 통해 면접을 보는 방식으로, 사람을 직접 마주 보지 않고 화면 안에서의 얼굴만을 볼 수 있기 때문에 대면 면접보다 면접자의 얼굴 안에서의 비언어적 표현이 더욱 중요한 평가 요소가 된다. 또한, 대면 면접에서는 면접관이 준비한 공간에서 면접자가 가서 면접을 보지만, 비대면 면접에서는 면접자가 준비한 공간에서 화상 통화 어플리케이션을 통해 면접을 봐야 하기 때문에 면접자가 미리 준비하고 고려해야 할 변수 사항이 더욱 많다. 하지만 현재 이러한 상황에도 불구하고, 면접자가 비대면 면접을 준비하고 연습할 환경은 턱없이 부족하다.

본 논문에서는 이러한 문제 해결을 위해 면접자가 비대면 면접을 종합적으로 준비하고 연습할 수 있는 환경을 제공하는 시스템을 제안한다. 제안된 시스템은 면접자의 비언어적 표현에 대해서 피드백을 제공하여 온라인 화상 면접에 도움을 주는 것을 목표로 한다. 면접자의 얼굴 위치, 비언어적 표현인 눈의 시선 처리, 그리고 입술의 움직임을 통해 면접자의 태도를 분석하고, 그와 관련한 피드백을 제공하여 면접자가 간단하게 자신의 면접 태도에 대한 피드백을 얻을 수 있다.

2.1 Online System

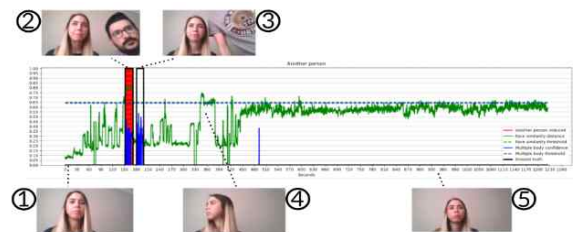


그림 1. 온라인에서의 부정행위 감지 연구

최근 언택트 시대에 접어들면서 온라인 시험 및 수업과 같은 온라인 시스템과 관련된 연구가 증가하고 있다. 특히 그림 1과 같이 얼굴을 인식하여 온라인상에서의 부정행위를 감지하거나 집중도를 판단하는 연구들의 중요성이 강조되고 있다. 화상 면접과 관련된 연구는 아직 부족한 상황이다.

2.2 Intersection over Union (IoU)

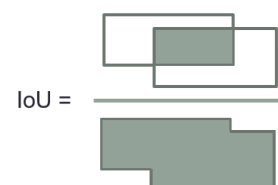


그림 2. Intersection over Union

11. 관련 연구

IOU는 Object Detection에서 두 물체의 위치가

얼마나 일치하는지를 판단하기 위해 사용된다. 그림 2와 같이 두 영역의 전체 넓이에 대한 두 영역의 겹쳐진 교차 영역의 넓이의 비율을 구하여 계산한다. 본 논문에서는 얼굴의 올바른 위치를 계산하는 과정에서 IoU 알고리즘을 적용하였다.

2.3 Gaze Estimation

Gaze Estimation은 사람의 눈을 통해 비언어적 표현인 시선을 추적하여 시선이 향하는 방향을 예측하는 분야이다. 동공을 감지하거나 head pose의 각도 계산식을 이용하는 등의 다양한 방법이 존재하며, 여러 분야에서의 응용이 가능하다. 본 논문에서는 2D로 눈에서 동공을 감지하여 시선의 수평 및 수직 좌표를 통해 시선을 예측하는 방법을 사용한다.

III. 본론

3.1 올바른 화상 면접

필요한 알고리즘을 구현하기에 앞서 먼저 올바른 화상 면접의 상황을 정의하였다. 본 논문에서는 시선 및 표정 처리와 화면 각도의 세 가지 측면에서의 중요성을 고려하였다. 첫 번째로 시선을 이리저리 돌리는 행동은 부정적인 요인이 되지만, 카메라 렌즈를 정면으로 바라보는 것은 면접에 집중하고 있다는 인상을 줄 수 있다. 두 번째로 답변을 하지 않는 상황에서는 무표정보다는 약간의 미소를 띄는 것이 중요하다. 마지막으로 카메라의 각도 조절을 통해 화면상에서 면접자의 얼굴이 한쪽으로 치우치지 않고, 중앙에 위치하는 것이 중요하다.



그림 3. 올바른 화상 면접의 예시

그림 3과 같이 면접자의 얼굴이 중앙에 위치하며 카메라의 정면을 주시하고 약간의 미소를 띄우는 경우를 온라인 화상 면접의 올바른 상황으로 정의하였다. 따라서 이를 기준으로 면접자가 면접관에게 긍정적인 인상을 줄 수 있도록 화상 면접 연습을 돕는 시스템을 구현하였다.

3.2 설계 블록도

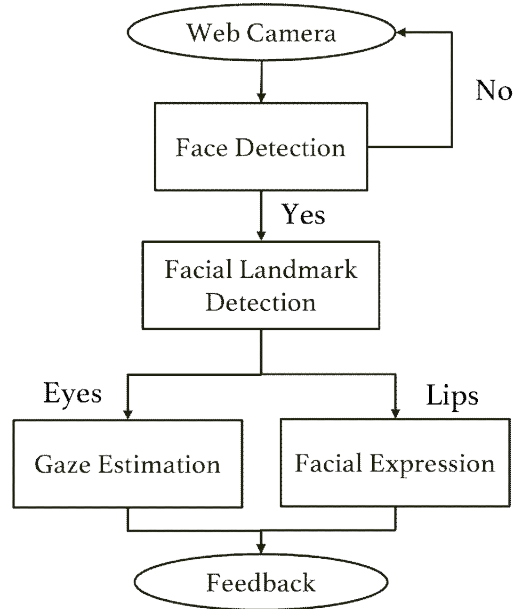


그림 4. 시스템 설계 블록도

본 논문에서 제시하는 시스템의 설계 블록도는 그림 4와 같다. 먼저 웹 카메라를 통해 들어오는 영상에서 frame 단위로 얼굴을 찾는 과정을 진행한다. 얼굴의 위치가 올바르다면 Dlib를 통해 이목구비의 landmark를 추출한다. 추출한 landmark 중 일부를 이용하여 시선 처리를 분석하는 Gaze Estimation과 입술 움직임에 따른 정보를 분석하는 Facial Expression 과정을 진행한다. 분석 과정이 모두 끝나면 최종 피드백을 면접자에게 제공한다.

3.3 Face Detection

화상 면접을 진행할 때 면접자의 얼굴은 한 방향으로 치우치지 않고 중앙에 위치해야 하므로 올바른 각도 조절이 필요하다. 따라서 화면상에서 적절한 얼굴의 위치를 그림 5와 같이 영상의 가로 길이에서 1/3지점부터 2/3지점까지, 그리고 영상의 세로 길이에서 1/4 지점에서 3/4 지점까지로 설정하였다.

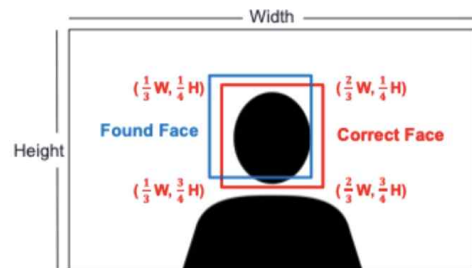


그림 5. Correct Face의 위치

이렇게 설정한 적절한 얼굴 위치와 Dlib 라이브러리에서 제공하는 pre-trained Face Detection model로 찾은 면접자의 얼굴 위치를 비교한다. 적절한 얼굴과 찾은 얼굴의 겹치는 영역이 찾은 얼굴에 대해 80% 이상인 경우는 적절한 경우로, 80% 미만인 경우는 잘못된 경우로 판단한다. 면접자의

굴 위치가 잘못된 위치로 판단된 경우, 그 다음 과정은 진행되지 않으며 면접자에게 경고 메시지를 출력한다.

3.4 Gaze Estimation

Gaze Tracking 오픈소스 프로젝트를 활용하여 Gaze Estimation 과정을 진행하였다. 면접자의 시선을 추적하기 위해서는 우선 눈에 대한 이미지를 얻어 동공을 추출해야 한다. Dlib 라이브러리를 이용해 얻은 Eye Landmark 값들을 통해 그림 6과 같이 눈에 대한 정보를 결정할 수 있다.



그림 6. Eye landmark에 따른 눈 정보

왼쪽 눈을 기준으로 했을 때, Eye landmark 중 x의 값이 가장 작은 좌표는 눈의 왼쪽(left)이 되고, x의 값이 가장 큰 좌표는 눈의 오른쪽(right)이 된다. 또한, 눈의 위쪽에 있는 2개 좌표의 평균 좌표가 눈의 위(top)가 되고, 아래쪽에 있는 2개 좌표의 평균 좌표가 눈의 아래(bottom)가 된다. 이렇게 얻은 눈의 4개의 좌표에 일정 여유를 두고 눈 이미지(Eye image)를 얻을 수 있다.



그림 7. Image Processing 진행 과정

눈 이미지에서 픽셀의 밝기 차이를 이용하여 검정색 부분을 추출하는 Image Processing 과정을 그림 7과 같이 진행한다. 눈의 검정색 부분인 동공 부분만을 추출하여 동공 프레임(Iris frame)을 구한다. 이 동공 프레임에 대한 Contour Detection을 진행하고, 추출한 Contour 내의 무게중심을 구하면 동공의 중심 위치가 된다. 이를 통해 면접자의 동공 중심 위치와 눈의 위치를 비교하여 시선을 추적할 수 있다.

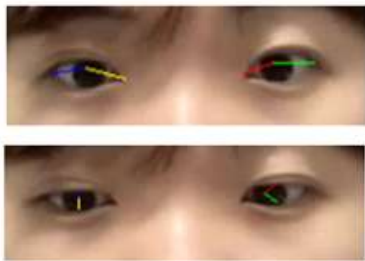


그림 8. 시선 추적을 위한 거리

면접자의 동공 중심과 눈 이미지에서 얻은 정보들에 대한 유클리드 거리 계산으로 시선을 추적한다. 시선의 수평 방향의 경우, 두 눈에 대해서 동

공의 중심과 눈의 왼쪽과 오른쪽에 대한 거리를 계산하여 각 눈에 대해 비율을 구하고, 각 비율들의 평균을 통해 눈의 중심을 기준으로 시선이 왼쪽인지 오른쪽인지를 구분한다. 이와 동일하게 시선의 수직 방향의 경우도 두 눈에 대해서 동공의 중심과 눈의 위와 아래에 대한 거리를 계산해 얻은 각 비율들의 평균을 통해 시선이 위인지 아래인지를 구분한다. 이러한 과정을 통해 시선이 향하는 방향을 상하좌우의 총 4가지 경우로 예측한다.

3.4.1 Image Processing

Gaze Estimation 중 픽셀의 밝기 차이를 이용하여 검은색 부분인 동공 부분을 찾는 과정이다. 시선의 위치 판단을 위해서는 눈 이미지에서 동공을 찾는 과정이 반드시 선행되어야 한다. 기존 오픈소스 알고리즘의 경우 이 과정에서 동공을 찾지 못해 시선 판단이 불가능한 경우가 종종 있었다. 이를 개선하기 위해 이 과정 중 찾은 객체의 크기를 감소하는 erode 연산의 parameter를 조정하였고, 이를 통해 객체의 크기 감소를 최소화하여 1.2% 성능을 개선하였다.

표 1. 알고리즘 개선 전과 후의 정확도

	개선 전	개선 후
전체 frame 개수	450개	450개
동공 찾은 개수	425개	430개
정확도	94.4%	95.6%

3.4.2 Blinking

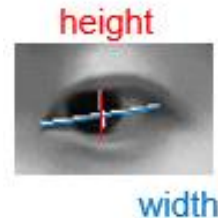


그림 9. 눈의 가로와 세로 길이

Eye landmark를 통해 구한 눈에 대한 정보인 눈의 왼쪽과 오른쪽을 연결하여 눈의 가로 길이를, 그리고 눈의 위와 아래를 연결하여 눈의 세로 길이를 구한다. 이렇게 구한 눈의 가로 길이와 세로 길이의 비율을 통해 면접자가 눈을 감았는지에 대한 상태를 판단한다. 눈을 감은 상태를 판단하는 비율식은 기존 오픈소스 알고리즘에서의 식이 아닌 일반적으로 사용되는 EAR (Eyes Aspect Ratio) 식을 활용하여 눈을 감은 상태를 판단하도록 하였다.

$$EAR = \frac{\|p2 - p6\| + \|p3 - p5\|}{2 \|p1 - p4\|}$$

그림 10. Drowsy 판단을 위한 EAR

기존 오픈소스 알고리즘의 경우 눈을 감아서 동공의 좌표를 못 찾는 경우, 감은 눈의 상태를 판단하는 것이 불가능했다. 그렇기 때문에 동공의 좌표가 없더라도 눈의 가로 길이와 세로 길이를 구할 수 있다면 상태 판단이 가능하도록 알고리즘을 개선하였다.

3.5 Facial Expression

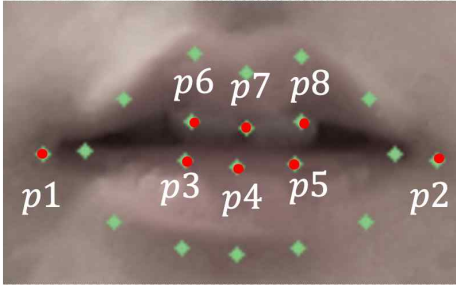


그림 11. 8개의 Mouth Landmark

Mouth Landmark를 활용해 면접자의 표정과 말하는 여부를 판단하는 과정이다. Dlib를 통해 추출한 총 68개의 Landmark 특징점 중 그림 11과 같이 필요한 8개의 점을 활용하였다. 8개의 Landmark를 통해 Facial Expression 과정을 수행하기 위한 LCR, LSR 계산식을 정의하였다.

$$LCR = \frac{\|p3 - p6\| + \|p4 - p7\| + \|p5 - p8\|}{3 \|p1 - p2\|} \quad (1)$$

먼저 눈의 blinking을 판단하는 EAR 계산식과 유사하게 말하고 있는지를 판단하기 위한 LCR(Lips Close Rate)를 수식 (1)과 같이 정의하였다. 입의 양끝 점과 가운데의 점들을 이용하여 면접자가 말을 하고 있는지 아닌지를 판단한다. 8개의 좌표들 중 p1과 p2를 연결하여 입의 가로 길이를 구하고, p3~p8을 연결하여 세로 길이를 구한다. 이때 LCR이 2.5보다 작을 때 말을 하고 있지 않는 close 상태로 판단한다.

$$LSR = \frac{2(p3 + p4 + p5)}{3(p1 + p2)} \quad (2)$$

또한 입의 양쪽 좌표와 중심 좌표를 비교하여 면접자가 웃고 있는지를 판단하는 LSR(Lip Smile Ratio)을 수식 (2)와 같이 정의하였다. 한다. 입의 양끝 점인 p1, p2와 중간 점인 p3~p5의 y 좌표의 평균 값을 통해 면접자가 웃고 있는지 아닌지를 판단한다. 이때 LSR의 비율이 0.99 이상일 때 웃고 있는 상태로 판단된다. OpenCV의 좌표계는 일반적인 좌표계와 달리 y좌표가 클수록 아래에 존재하기 때문에 비율이 작을 때가 아닌 클 때 입꼬리가 올라가 있다고 판단된다.

IV. 결과 및 분석

4.1. Dataset

구현한 알고리즘의 성능 평가를 위해 실험용 Dataset을 준비하였다. 먼저 본 논문에서 정의한 올바른 화상 면접의 상황과 올바르지 않은 상황의

2가지 경우를 가정하였다. 총 10명의 사람을 대상으로 하여 해상도 1920x1080의 2분 미만 영상을 15fps로 촬영하였다. 또한, 실제 면접과 유사한 상황을 연출하기 위해 화상 회의 프로그램인 zoom에서의 영상을 촬영하였다.

4.2 Test 결과

표 2. Threshold 별 알고리즘 정확도

Threshold	Blinking	Threshold	Smile
0.21	96.2%	0.98	74.26%
0.23	96.7%	0.99	87.57%
0.25	92.8%	1.0	70.12%

구현한 알고리즘의 성능 평가와 threshold의 결정을 위해 수집한 20개의 dataset을 바탕으로 정확도를 구하였다. 정확도는 전체 데이터 수 대비 올바르게 예측된 데이터의 수의 비율로 계산하였다.

실험 결과, Blinking Ratio의 정확도는 threshold가 0.23일 때 가장 높은 96.7%의 값을 가지며, Smile Ratio의 정확도는 threshold가 0.99일 때 가장 높은 87.57%의 값을 가진다. 따라서 각각의 알고리즘에 가장 높은 정확도를 갖는 threshold 값을 기준으로 설정하였다.

4.3 피드백 결과

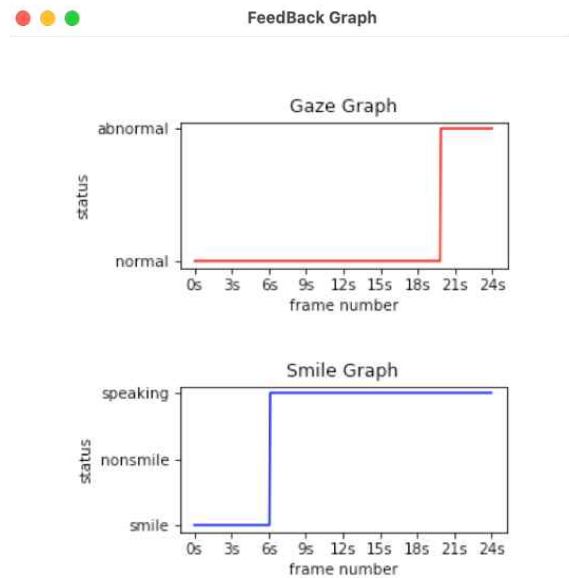


그림 12. GUI 상에서 출력되는 그래프

면접자의 영상 분석이 끝난 후 제공되는 피드백 그래프는 그림 12와 같다. Matplotlib 라이브러리를 활용하여 시선 처리와 표정에 대한 두 가지 그래프를 피드백으로 출력한다. 영상에서의 시간이 가로축에 해당되고, 각 상태가 세로축에 해당되므로 전체 영상에서 해당하는 시간에 대한 상태를 한눈에 파악할 수 있다.

이때 출력되는 그래프의 세로축은 단순한 상태 판단이 아닌 누적값을 고려하여 계산된다. 즉, 올바르게 않은 상태가 일정 시간 이상 지속되는 경우

에만 실제로 올바르게 맞다고 판단한다. 마찬가지로 반복되던 중 정상적인 경우가 잠깐이라도 존재하는 경우 상태가 다시 변할 것이다. 이를 방지하기 위해 마찬가지로 정상적인 경우도 일정 시간 이상 반복되는 경우만 변화시키도록 하였다. 본 논문의 경우, 1초 이상 올바르게 맞은 경우가 반복되는 경우 실제로 올바르게 맞다고 판단하였고, 다시 0.5초 이상 올바른 경우가 반복되는 경우 상태를 변화시켜 출력 그래프에 반영하였다.

4.4 시스템 UI



그림 13. GUI 프로그램 구현 결과

사용자에게 편의성을 제공하기 위해 Tkinter 라이브러리를 활용하여 전체적인 GUI 프로그램을 구현하였다. 제공되는 프로그램은 그림 13과 같이 구성된다. 메뉴 기능을 통해 미리 촬영된 영상을 업로드하거나 실시간에서의 촬영을 선택할 수 있다. 모드를 선택 후 분석을 시작하면 전체 영상에서의 분석을 진행하고, 완료된 피드백 그래프와 영상이 저장된다.

또한, 실시간 분석에서 그림 13과 같이 면접자의 얼굴이 올바른 위치에 들어오지 않은 경우, 화면상에 경고문을 출력하며 분석을 진행하지 않는다. 이때 분석이 시작되면 사용자의 시선 처리와 표정에 대한 상태는 좌측 하단에 실시간으로 출력되며, 누적값이 고려된 경고문의 경우 우측 상단에 출력된다.

V. 결론

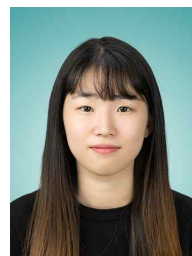
본 논문에서는 면접자의 비언어적 표현을 분석하여 온라인 화상 면접을 연습할 수 있는 시스템을 제시한다. Facial Landmark를 통해 시선 처리와 표정을 분석하고, 이에 대한 피드백 결과를 면접자에

게 제공한다. 이는 단순한 상태 판단이 아닌 시간적 측면이 고려되므로 실질적인 영상에서의 활용이 가능하다. 또한 전체 과정이 사용자에게 편리한 UI로 제공된다는 점에 의의가 있다. 따라서 면접자는 제공되는 피드백 결과를 통해 진행 과정에서 본인의 습관을 발견하고 이를 개선하여 면접관에게 긍정적인 인상을 남길 수 있다.

다만, 본 논문에서 제시하는 시스템은 다음과 같은 한계가 존재한다. 먼저 음성 데이터가 아닌 시각적 데이터만을 활용하기 때문에 이를 보완한다면 더 다양한 피드백 결과를 기대할 수 있다. 또한, 구현된 UI 상에서의 분석 시간이 오래 걸리기 때문에 multi-thread 등을 통한 해결이 필요하다. 이러한 한계점들을 개선한다면 면접자에게 더욱 도움이 되는 실용적인 연습 시스템을 제공할 수 있을 것이다. 더 나아가 언택트 시대에 접어들어 현재, 온라인 수업 및 화상 회의와 같은 여러 온라인 시스템에서 본 논문에서 제시하는 알고리즘의 적용 효과를 볼 수 있으리라 기대한다.

참고문헌

- [1] <https://awesomeopensource.com/project/antoine/elame/GazeTracking>
- [2] "Cheating Detection Pipeline for Online Interviews and Exams", Azmi Can Ozgen, 2021, IEEE
- [3] "A Visual Analytics Approach to Facilitate the Proctoring of Online Exams", HaotianLi, Min Xu, Yong Wang, Huan Wei, HuaminQu, 2021, arXiv:2101.07990
- [4] 'HLR-Net: A Hybrid Lip-Reading Model Based on Deep Convolutional Neural Networks', AmanyM. Sarhanand Nada M. Elshennawyand Dina M. Ibrahim,Cmc-computers Materials W& Continua, 2021, 68, 1531-1549
- [5] "Real-Time Eye Blink Detection using Facial Landmarks", TerezeSoukpovaand Jan Cech, Computer Science. (2016)
- [6] MeeyeonOh, YoosooJeong, Kil-HoumPark, "Driver Drowsiness Detection Algorithm based on Facial Features", Journal of Korea Multimedia Society Vol. 19, No. 11, November 2016 (pp. 1852 - 1861)
- [7] JeaheeLee, GoomanPark, "얼굴인식 기반 비대면 온라인 강의학습보조 시스템", 한국방송미디어 공학회 2020 추계학술대회 (pp. 346 - 348), 2020.11



강 유 립

2017 ~ 현재 인하대학교 공과대학 정보통신공학과 학사 과정 재학 중

관심 분야는 소프트웨어 개발, 웹 개발, 데이터베이스



박 지 은

2017 ~ 현재 인하대학교 공과대
학 정보통신공학과 학사 과정 재
학 중

관심 분야는 소프트웨어 개발,
데이터베이스, 영상 처리