360 카메라를 이용한 디지털 트윈 강의실

유현태*, 김진호*, 김유성*, 박인규*
*인하대학교 정보통신공학과
helios1127@naver.com, jinho9610@naver.com, bnb407@naver.com, pik@inha.ac.kr

Digital Twin Classroom using 360 Camera

Hyeontae Yoo*, Jinho Kim*, Yoosung Kim*, Inkyu Park*
* Department of Information and Communication Engineering, Inha University

요 약

본 논문에서는 딥러닝 얼굴 인식을 이용하여 실시간 360 공간 Classroom 과 실시간을 기반으로 한 가상 360 공간 Classroom 을 제안한다. MTCNN 을 이용한 얼굴 검출 및 Inception Resnet V1 모델을 이용한 딥러닝 기법을 통해 얼굴인식을 진행하고 HSV 색공간 기반의 화자 판별, 아바타 Rendering, 출석 체크 등을 진행한다. 이후 시각화를 위해 제작한 Web UI/UX 를 통해 사용자에게 현실과 가상 공간을 넘나드는 Twin Classroom 을 제공한다. 따라서 사용자는 새로운 화상 교육 플랫폼에서 보다 개선되고 생동감 있는 Classroom 에서 교육을 받을 수 있다.

1. 서론

COVID-19 팬데믹 선포 이후, 사회 활동이 비대면으로 급격히 전환되고 있다. 그러나 사용자가 급증하면서 사용자들이 플랫폼 이용시에 호소하는 문제점들도 함께 증가하고 있다.

먼저 비대면 수업시에 참여 인원이 증가함에 따라 전통적인 출결 확인 방식으로 출결 처리를 하기에는 모든 이용자간 커뮤니케이션이 원활하지 않을 수 있다. 또한 참여자는 개인 단말 카메라를 이용하므로, 모든 이용자를 한 번에 화면에 담을 수 없다는 문제가 있어서 현장감의 부재와 같은 문제가 발생할 수 있다. 또한, 하나의 영상에 여러 사용자를 송출하므로 발표시에 실제 발화자 파악의 어려움과 같은 문제가 발생하며 발화자가 여러 명인 경우, 음성 충돌 등의 문제 또한 발생할 수 있다. 이렇게 다양한 문제점들이 발생하기 때문에 기존의 화상 플랫폼을 그대로 사용하기에는 불편함이 동반될 수밖에 없다.

본 논문에서는 위에서 기술한 문제점들을 해결하기 위해 구체적으로 5 가지의 기법을 설명하며 최종적으로 이 기술들이 모두 접목된 Digital Twin Classroom 을 제시한다.

본 논문의 구성은 다음과 같다. 2 절에서는 각 기법의 사용 목적과 상세 구현 내용, 결과, 사용시에 추구할 수 있는 이익을 설명하고 3 절에서는 이러한 기법들의 통합 구성도를 설명한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 시스템

2-1. 딥러닝을 이용한 얼굴 인식

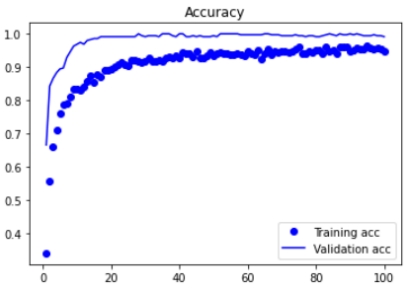
Pretrained 된 VGGFace2 모델을 사용하여 GAP(Global Average Pooling Layer) 전까지의 네트워크를 freeze 하여 학습이 되지 않도록 하고, 팀원 3 명에 대하여 3 개의 클래스로 분류될 수 있도록 전이 학습을 진행한다 [1].

학습에 필요한 Dataset 확보를 위해 팀원 3 명에 대해 각자 얼굴 20 장을 촬영하고, 눈을 잇는 두 직선을 기준으로 얼굴 수평을 맞춘 후, 좌우로 0~10 도 중 임의의 각도 회전을 주는 방식으로 data augmentation 을 진행한다. MTCNN 을 이용하여 각 사진에서 얼굴을 검출한다. 이때, 얼굴의 중심을 기준으로 정사각형으로 추출한다. 마지막으로 224x224 size 로 얼굴을 확대하여 기본 Dataset 을 구축한다. 학습 진행 후 각각의 사람에 대해 8 장씩 총 40 장의 test 용 Data set 을 생성 후 그 성능을 측정하였을 때 95% 정도의 성능을 보였다. 하지만 같은 방식으로 360 영상에서 얼굴 검출을 진행하여 모델을 통해 예측을 하였을 때 아주 좋지 못한 성능을 얻을 수 있었다. 그 원인으로 가로의 길이가 6K 인 360 영상에서 얼굴 검출시에 추출된 얼굴 크기가 약 30x30 정도를 보여 이를 학습시에 224x224 로 resize 하는 과정에서 극심한 화질의 저하가 발생하기 때문이라 판단한다.

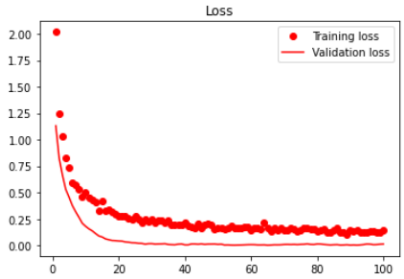
이에 학습용 Dataset 의 해상도를 축소시킬 필요성을 느꼈다. 먼저 80x80 으로 진행하였을 때 좋지 못한 성능을 보였고, 100x100 으로 약간 더 확대하여 Dataset 을 구축하고 성능을 확인해 보았을 때 꽤 괜찮은 성능을 보였다. 따라서 이후

360 카메라를 이용하여 개개인 별 약 15 초의 동영상을 촬영하여 임의의 프레임을 선택하여 얼굴을 검출하고 얼굴 size 를 100x100 으로 resize 하는 방식으로 Dataset 을 구축한다.

최종적으로 Data augmentation 후 학습용 data 280 장, validation 용 data 70 장을 생성하였고 학습용 Data set Batch Size 64, Validation Data set Batch Size 32, Epoch 100 으로 진행하여 다음과 같은 러닝 커브를 얻을 수 있다.



<그림 1. Transfer Learning model 의 Accuracy>

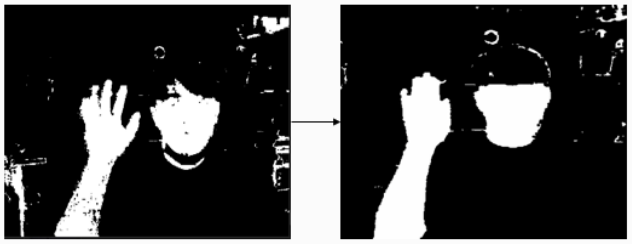


<그림 2. Transfer Learning model 의 Loss>

위의 모델을 이용하여 360 카메라에서 검출된 얼굴에 대해 테스트를 진행하였을 때 93.5%까지 성능을 향상시킬 수 있었다.

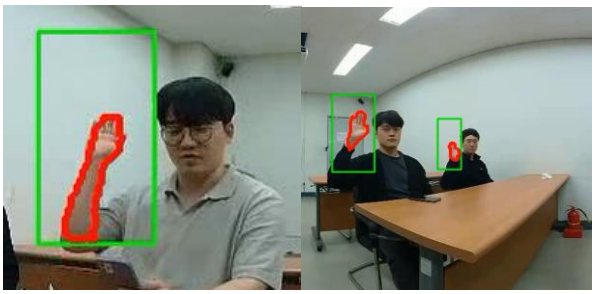
2-2. HSV 색공간을 이용한 손 검출

HSV 색공간을 이용한 손 검출은 입력 영상인 360 영상내 HSV 색공간에서의 일정 구역안에서 살색을 감지해 손을 검출한다. 먼저, HSV 색공간에서 살색을 흰색으로, 나머지 색공간은 검은색으로 변환해 이진 영상을 만들었다. 이때, 손과 팔이 정확하게 검출이 되지 않는 것을 우려해 그림 3 와 같이 Morphology 연산을 적용시켜 검출성능을 향상시켰다.



<그림 3. Morphology 연산 전, 후>

그림 4 와 같이 MTCNN 모델 기반으로 검출한 얼굴 좌표의 근처 구역에 일정한 직사각형의 영역을 두어 해당 영역에서의 살색을 검출하게 된다. 해당 영역에서 기존의 살색 영역의 컨투어 면적을 파악해 일정 값 이상이 될 경우 해당 인원이 손을 들었음을 감지한다.



<그림 4. 360 영상에서의 손 검출>

2-3. Blender, Three.js 를 이용한 가상 공간 렌더링

[2]를 이용한 실제 2D 영상을 3D obj 로의 복원한 것을 Blender 를 통해 렌더링 및 텍스처 매핑을 진행한다. 각각의 3D obj 와 사람 상체를 결합해서 사람 obj 를 렌더링한다 [2, 3]. 또한, 책상 및 의자 렌더링 후 360 카메라로 취득한 실제 교실 영상을 바탕으로 큐브 맵을 만들어 교실 공간을 만든다. 각각의 만들어진 obj 를 Web 에 렌더링 하기 위해 JavaScript 의 라이브러리인 Three.js 를 이용해서 웹에 렌더링한다. 또한, Web 에 렌더링 된 가상 공간에서 마우스 및 키보드 입력으로 자유로운 시점이동이 가능하다. 또한, 출석 확인 및 화자 판별을 검출하기 위해 렌더링 된 사람의 우측 상단에 이름표를 렌더링 한다. 자유로운 시점이동에 따라 이름표 또한 이동하는 시점에 맞추어 방향이 바뀌는 Billboard 기능을 추가해 어떠한 시점에서도 이름표를 정면에서 보는 것과 같은 효과를 준다.

2-4. Flask, JSON 를 이용한 웹 송출

Insta360 One R 카메라를 이용한 실시간 스트리밍은 Insta360 앱 내에서 YouTube, Facebook, Kawai 을 거치도록 송출이 제한된다. 그 중 우리는 YouTube 를 거쳐 360 영상의 frame 을 받아온다. Flask 프레임 워크를 이용해 얼굴 인식, 손을 검출한 데이터를 SSE 통신을 이용한 JSON 스트리밍을 통해 Web 에 송출한다. 그림 5 과 같이 얼굴 인식이 될 경우에 Three.js 를 이용해 렌더링한 이름표의 색이 빨간색색에서 파란색으로 변경이 된다. 그리고 손을 든 동작을 검출할 경우에 얼굴 인식의 경우와 마찬가지로 “Hand up”이라는 문구의 텍스트를 이름표의 위에 렌더링한다.

<그림 5. 교실 내에 없는 경우(좌측), 교실 내에서 손을 든 경우(우측)>

2-5. 전체 시각화

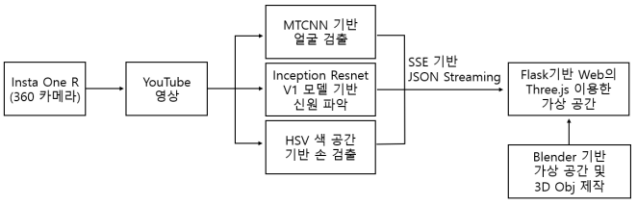
Web UI/UX 를 통해 사용자의 편의에 따라 실제 360 카메라로 송출되고 있는 영상을 볼 수 있는 Real Spatial 과 실제 공간을 반영하여 만든 Virtual Spatial 을 선택할 수 있는 Button 을 구현하였다. 또한, 오른쪽 Nav Bar 를 이용하여 출결 및 현재 재실 여부를 알려주는 UI 를 구현한다. 360 영상 송출시에 YouTube Live 송출 시스템을 이용해서 결과적으로 그림 6 과 같은 Real Spatial UI 를 얻는다.

<그림 6. Real Spatial Web 화면>

하지만 일반 노트북 화면 크기를 기준으로 송출한 360 영상을 resize 하는 과정에서 약간의 화질 저하가 일어난다. 이에 사용자의 설정에 따라 Virtual Spatial Button 을 통해 가상 공간의 좀더 매끄러운 환경에서 수업을 들을 수 있다. AR 기반의 가상 환경에서는 실제 현장에서 참여하고 있는 학생들의 위치를 파악하여 아바타를 렌더링 해줌으로써 다소 떨어질 수 있는 생동감을 보완해줄 수 있다.

3. 구성도

Insta One R 360 카메라를 이용해 YouTube 플랫폼에서 Live Streaming 한다. 360 카메라를 통해 각각의 프레임을 MTCNN 기반 얼굴 검출, Inception Resnet V1 모델 기반 신원 파악, HSV 색공간 기반 손 검출을 하게 된다. 각각의 정보를 SSE 기반 JSON Streaming 을 통해 Flask 로 제작된 Web 에 데이터를 전송한다. Web 내에 Blender 로 생성된 Obj 와 Three.js 기반으로 구축된 가상공간에서 다양한 기능을 구현한다.



<그림 7. Digital Twin Classroom 구성도>

4. 결론

본 논문에서는 위 기술한 모든 기법들을 통합하여 만든 Digital Twin Classroom 은 현장에 참가할 수 없는 학생들을 대상으로 한다. 360 카메라 영상에서의 얼굴 인식을 기반으로 출석 체크, 실시간 재실 여부 등을 한 눈에 파악할 수 있고, 다른 학생이 발표를 하고 있는지 등을 파악할 수 있다. 또한, 가상 공간에서는 자유로운 시점 이동을 통해 교육환경을 확인할 수 있다. 본 논문에서 설계하고 구현한 프로그램은 같은 공간에 있지 않는 학생들에게 생동감 있는 모습을 제공할 것이라 예상한다.

5. 감사의 글

본 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2020-0-01389, 인공지능융합연구센터지원(인하대학교)).

참고문헌

[1] Christian Szegedy, Google Inc, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”, 2016

[2] In Kyu Park, hui Zhang, Vladimir Vezhnevets, “Image-Based 3D Face Modeling System”, 2005

[3] J. Lee, 3D Face and Head Reconstruction with Geometric Details using Deep Neural Networks, Master’s Thesis of Inha University, Incheon, Korea, 2020.