화상회의 서비스를 위한 GAN 기반 아바타 생성 및 애니메이션 구현 기술

문지언[1], 김지윤[1], 박지혜[1], 안효원[1], 이경미[1]
[1]덕성여자대학교 컴퓨터공학과
answldjs1836@duksung.ac.kr, jiyoon0417@duksung.ac.kr,
sia20181011@duksung.ac.kr, hw971204@duksung.ac.kr, kmlee@duksung.ac.kr

GAN-based avatar generation and animation for video conferencing service

Ji-Eun Moon[1], Ji-Yun Kim[1], Ji-Hye Park[1], Hyo-Won Ahn[1], Kyoung-Mi Lee[1]
[1]Dept. of Computer Science, Duksung Women's University

요 약

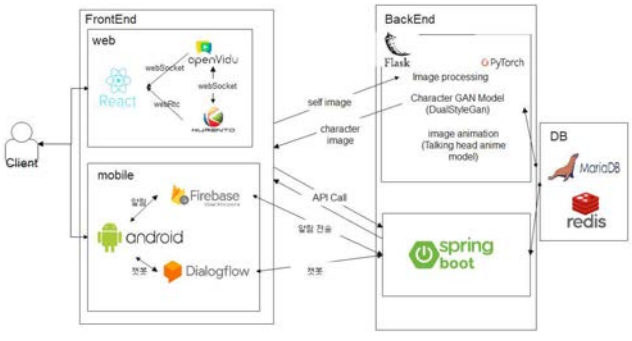
코로나19 이후 화상회의 빈도가 높아지면서 줌 피로라는 신조어가 등장할 만큼 상대방을 가까이 마주하며 회의를 진행하는 것이 사람들의 피로도를 상승시키고 있다. 본 논문에서는 얼굴 합성과 이미지 애니메이션을 이용한 아바타를 통해 사용자가 화상회의에 참가할 수 있는 시스템을 제안한다. 사용자와 닮은 개성 있는 캐릭터는 실시간으로 사용자의 표정 및 움직임을 반영하여 화상회의에 적용될 수 있고 채팅과 커뮤니티에서 캐릭터의 이모티콘으로 감정을 표현할 수 있다.

1. 서론

코로나19 이후 늘어난 화상회의로 해외에서는 '줌 피로(zoom fatigue)[1]'라는 신조어가 등장했다. 줌 피로를 유발하는 원인 중 하나로 화상회의 화면에 상대방의 얼굴과 함께 뜬 자기 모습이 스트레스를 유발한다고 한다. 자기 모습이 아닌 아바타로 회의하게 된다면 더 나은 환경이 조성될 것이다.

따라서 본 연구에서는 자신만의 아바타를 만들어 화상회의를 하고 자신의 아바타 이모티콘를 활용하여 회의채팅 및 커뮤니티에서 활용할 수 있게 구현하였다.

2. 화상회의 시스템 아키텍쳐



(그림 1) 시스템 아키텍쳐

본 논문에서 제안하는 화상회의 서비스의 시스템 아키텍쳐는 그림 1과 같다. 아바타 생성 모델 DualStyleGan과 이모티콘, 이미지 애니메이션 생성 모델 Talking head anime model은 서버 flask api로 제공한다. 사용자, 일정, 게시글 api는 spring boot 서버를 통해 제공한다. 클라이언트는 웹과 모바일을 제공하여 사용자의 편리성을 보완한다. 웹은 react.js를 사용해 구현하며, openvidu를 활용해 화상회의 서버를 구성한다. 모바일은 Android를 사용해 구현한다. 휴대성이 편리하다는 모바일의 장점을 극대화하기 위해 일정 알림과 챗봇 기능을 추가한다. 알림은 FCM을 활용하고 챗봇은 Google DialogFlow를 활용해 구현한다.

3. 아바타 생성을 위한 GAN 기반 적용 모델 및 기술

3-1. 캐릭터 생성 적용모델 - DualStyleGan

DualStyleGan은 실제 얼굴에 새로운 캐릭터 얼굴의 스타일을 특성화하여 자연스러운 얼굴 합성을 만들어내는 모델이다[2]. DualStyleGan은 기존 StyleGan2에서 새로운 네트워크를 추가해 한 쌍의 스타일 네트워크를 사용하며, 다른 모델에 비해 데이터셋이 적어도 성능이 우수한 모델이다. 본 프로

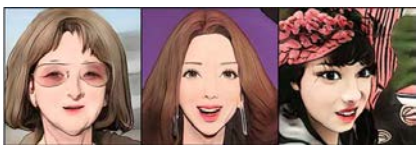
젝트는 DualStyleGan을 통해 실제 인물과 캐릭터를 매칭하여 나만의 아바타를 생성한다.

대부분의 기존 GAN 모델들이 서구적인 만화 캐릭터들을 대상으로 학습되었다는 점을 보완하기 위해 본 연구는 한국형 만화 캐릭터를 사용한다. '여신강림'은 네이버 웹툰 중 신드롬을 일으켰던 한 작품으로 Naver, Google 등 웹사이트에서 크롤링을 통해 데이터를 수집하였다.

캐릭터 생성을 위한 학습 단계는 크게 세 가지로 나뉜다. 첫 단계인 얼굴 디스타일화(destylization)는 예술적 초상화에서 사실적인 얼굴을 복구하는 것으로 고정된 얼굴-초상화 쌍을 형성하는 것을 목표로 한다(그림2).

(그림 2) 얼굴 디스타일화 결과

다음은 점진적 미세 조정 단계로, 목표 도메인의 색상, 얼굴 구조, 얼굴 스타일을 차례대로 기존 도메인에 전송한다(그림 3).



(그림 3) Fine-Tuning 결과

마지막 단계는 잠재 최적화와 샘플링으로, 조정된 모델로 얻은 이미지는 훨씬 스타일 이미지에 가까워진다.

그림 4는 '여신강림'의 만화 캐릭터를 이용하여 사용자와 닮은 최종 학습된 캐릭터 이미지다. 그림 4의 왼쪽 이미지는 사용자, 가운데 이미지는 점진적 전이학습을 실행한 결과, 오른쪽 이미지는 스타일이 전송되어 최종 학습된 결과이다.

(그림 4) 네이버 웹툰 '여신 강림' 최종 학습 결과

3-2. Talking head anime model

Talking head anime model은 한 장의 2D 이미지를 애니메이션과 같이 표정을 바꾸고 3D 움직임을 얻을 수 있는 GAN 모델로[3], 본 연구에서는 아바타 애니메이션 기능과 이모티콘 기능을 제공한다.

Talking head anime model은 눈썹, 눈, 입, 눈동자 크기 및 위치, 얼굴 회전의 값을 가진 6차원 포즈(pose) 벡터를 face morpher와 face rotator 네트워크로 전달해서 아바타 이미지를 변경한다. 먼저 face morpher는 눈과 입이 열리는 정도로 표정을 변화시킨다[4]. 식 (1)은 face morpher의 손실함수이다.

$$L_{fm} = E_{(I_r,p,I_e)\sim p_{data}}\left[\| I_e - G_{fm}(I_r,p) \|_1 \right] \quad (1)$$

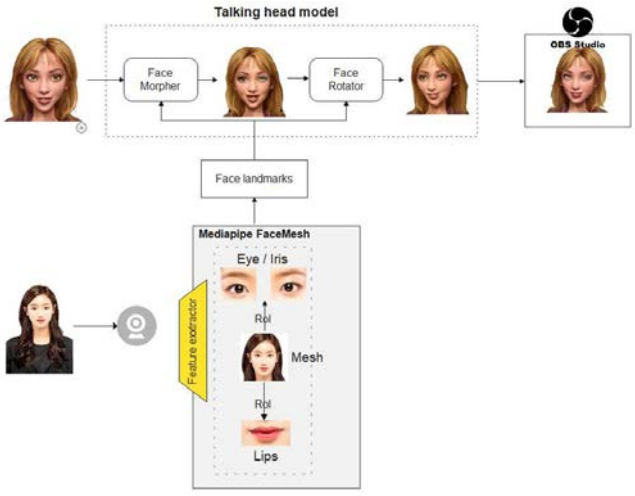
여기서 p_{data}는 훈련 데이터의 확률 분포를 나타낸다. (I_r,p,I_e)는 각각 I_r은 원본 이미지, p는 포즈 벡터, I_e는 변경 사항이다. G_{fm}은 face morpher 네트워크를 나타낸다.

다음 face rotator는 얼굴을 회전시킨다. 얼굴 회전은 2D가 아닌 3D 공간에서 이루어지므로 2D 이미지를 3D로 변환시키는 알고리즘이 필요하다. Face rotator는 Pumarola *et al.* [4]와 Zhou *et al.*[5]의 두 개의 뷰 합성 알고리즘의 결과값을 더하여 얼굴 회전을 계산한다. 식 (2)는 face rotator의 손실함수이다.

$$L_{fr}^{P} = \sum_{k=1}^{2} E_{(I_e,p,I_p)\sim P_{data}}\left[\| I_p - I_k \|_1 + \sum_{j=1}^{3} \lambda_j^{fr} \Phi_j(I_p,I_k) \right] \quad (2)$$

여기서 I_p는 포즈된 이미지이고 I_k는 Pumarola *et al.* [4]와 Zhou *et al.* [5] 네트워크를 통과한 이미지이다. $\Phi_j(I_p,I_k)$는 이미지 분류용 알고리즘인 VGG-16을 정의한다[6].

3-2-1 아바타 애니메이션



(그림 5) 아바타 애니메이션 흐름도

그림 5는 본 연구에서 구현한 아바타 애니메이션 흐름도를 보여준다. 먼저 MediaPipe를 이용하여 웹캠을 통해 사용자의 움직임을 실시간으로 얻어낸다. Mediapipe의 FaceMesh은 실시간으로 사용자의 얼굴의 특징을 추출할 수 있는 솔루션으로, 눈과 눈동자, 눈썹, 코, 입술, 그리고 얼굴 윤곽의 랜드마크를 추출하여 위치를 얻을 수 있다. 이렇게 추출한 얼굴 랜드마크 값을 Talking head anime model에 보낸다. Talking head anime model은 전달받은 랜드마크 값을 face morpher와 face rotator 네트워크로 전달하여 DualStyleGan을 통해 만들어진 고정된 포즈의 2D 아바타 이미지가 실시간으로 추출된 사용자의 얼굴 랜드마크 값에 따라 실시간으로 움직이는 모습을 보여줄 수 있다.

3-2-2 이모티콘

본 연구에서는 사용자의 의사 표현 전달을 위하여 이모티콘을 활용한다. 현재 이미지 애니메이션에서 사용하는 Talking head anime model이 실시간으로 눈, 코, 입, 귀의 좌표값을 가져오기 때문에 원본 이미지에서 각 좌푯값을 조절하여 다양한 표정의 이모티콘을 만들 수 있다. 고정된 표정의 2D 캐릭터를 다시 [기쁨, 슬픔, 화남, 윙크] 총 4가지의 표정을 가진 캐릭터로 변경, 생성시킬 수 있다(그림 6).

(그림 6) 이모티콘 구성도

먼저, 2D 캐릭터 이미지는 256×256의 크기로 RGBA 형식을 가지며 배경이 투명해야 한다. 이를 위해 이미지의 크기를 재조정하고, 컬러에 알파 채널을 추가한다. 얼굴 포즈 값에 각 표정에 따른 고정된 수치값을 적용한다. 포즈 값은 0~1 사이의 실수로 1의 값과 가까워질수록 뚜렷한 표정 변화가 나타난다. 본 연구는 정확한 표정을 보여주기 위하여 최댓값인 1을 적용한다. 예를 들어, [기쁨] 표정의 경우 눈과 입에 해당하는 포즈 벡터값을 조절하여 웃고 있는 눈과 올라간 입 모양을 표현하였다.

4. 결론

본 연구는 DualStyleGan과 Talking head anime model을 이용하여 자신과 닮은 한국형 만화 캐릭터를 생성하였고, 그 캐릭터를 이모티콘화 하여 화상회의에 활용할 수 있게 구현하였다. 아바타를 활용한 화상회의는 사생활을 보호해준다. 아바타라는 자신의 분신을 통해 평소 발표에 미숙한 사용자도 거리낌 없이 발표할 수 있는 환경이 조성되도록 도움을 준다. 또한 이모티콘을 활용하여 감정을 이미지화하여 전달해줌으로써 말로만 서술하는 것보다 자기 생각이나 감정을 더 효과적으로 전달할 수 있을 것이다.

본 연구는 향후 엔터테인먼트 업계 전반에서 활약하는 가상 인간(버추얼 휴먼)을 위한 가상 얼굴 디자인 분야, 버튜버(버추얼 유튜버)를 위한 가상 캐릭터 등 가상 캐릭터를 활용하는 다양한 플랫폼에서 활용할 수 있을 것이다.

Acknowledgments

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] 이은지, "MZ세대의 화상회의 피로감," *The Journal of the Convergence on Culture Technology,* Vol. 8, No.3, pp. 589-594, 2022.

[2] S. Yang, L. Jiang, Z. Liu, and C.C. Loy, “Pastiche master: Exemplar-based high-resolution portrait style transfer,” *In proceedings of the conference on CVPR*, p. 7693-7702, 2022.

[3] P. Khungurn, “Talking head anime from a single image 2: More expressive,”http://pkhungurn.github.io/talking-head-anime-2/, 2021.

[4] A. Pumarola, A. Agudo, A.M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, “GANimation: Anatomically-aware facial animation from a single image,” *In proceedings of the conference on ECCV*, pp. 835 - 851, 2018.

[5] E. Zhou, Z. Cao, and J. Sun, “GridFace: Face rectification via learning local homography transformations,” *In proceedings of the conference on ECCV*, pp. 3 - 20, 2018.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.