

사용자 아바타 생성과 애니메이션을 이용한 화상회의 시스템 개발

안 효 원¹ · 박 지 혜² · 김 지 윤² · 문 지 언² · 이 경 미^{3*}

¹고려대학교 컴퓨터학과 학부과정

²덕성여자대학교 컴퓨터공학과 학사

^{3*}덕성여자대학교 컴퓨터공학과 교수

Video Conferencing System Using User-Specific Avatar Generation and Animation

Hyo-Won Ahn¹ · Ji-Hye Park² · Ji-Yun Kim² · Ji-Eon Moon² · Kyoung-Mi Lee^{3*}

¹Bachelor's Course, Department of Computer Science, Korea University, Seoul 02841, Korea

²Bachelor's Degree, Department of Computer Science, Duksung Women's University, Seoul 01369, Korea

^{3*}Professor, Department of Computer Science, Duksung Women's University, Seoul 01369, Korea

[요 약]

최근 화상회의 시스템은 사용자가 화면에서 비춰진 자신의 모습을 대신할 수 있도록 아바타 사용을 적극 검토하고 있다. 그러나 기존 화상회의 시스템은 미리 만들어 놓은 동물 캐릭터를 아바타로 제공하기 때문에 사용자와 아바타 사이의 갭이 크다는 단점이 있다. 본 논문에서는 사용자별 아바타 생성과 이미지 애니메이션을 이용하여 현실감을 높인 화상회의 시스템을 제안한다. 제안하는 시스템은 DualStyleGAN을 이용하여 한국 웹툰 ‘여신강림’ 만화 캐릭터를 학습시켜 사용자를 닮은 한국형 캐릭터 스타일의 아바타를 생성한다. Talking head anim 모델을 이용하여 생성된 아바타의 눈, 눈썹, 입, 얼굴 등을 변형하는 방식으로 아바타가 실시간으로 사용자가 말하는 모습을 따라 움직이게 하고 아바타의 얼굴에 감정을 표현하는 이모티콘들을 생성한다. 또한 회의 일정 관리를 도우는 챗봇을 추가하여 화상회의 사용자의 편의성을 높인다.

[Abstract]

Recently, video conferencing systems have actively employed an avatar that can replace a user's face on the screen. However, existing video conferencing systems have a disadvantage in that the gap between the user and avatar is large because the avatar is a pre-made animal character. In this paper, we propose a video conferencing system that enhances reality through user-specific avatar generation and image animation. The proposed system uses DualStyleGAN to learn the Korean webtoon “True Beauty” cartoon characters and to generate a Korean character-style avatar that resembles the user. Using the Talking head anim model, the proposed system transforms the eyes, eyebrows, mouth, face, etc. of the generated avatar so that the avatar animates along with the user's speaking in real time and creates emoticons that express emotions on the avatar's face.

색인어 : 화상회의 시스템, 아바타, 생산적 적대 신경망, 애니메이션, 감정

Keyword : Video Conferencing, Avatar, Generative Adversarial Network, Animation, Emotion

<http://dx.doi.org/10.9728/dcs.2023.24.5.1061>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 04 March 2023; **Revised** 07 April 2023

Accepted 25 April 2023

***Corresponding Author; Kyoung-Mi Lee**

Tel: +82-2-901-8348

E-mail: kmlee@duksung.ac.kr

1. 서론

코로나19 대유행이 발생하면서 바이러스의 확산을 막고자 시행되었던 사회적 거리두기의 영향으로 비대면 서비스의 이용이 급격하게 증가하였다. 특히 비대면 미팅을 위한 화상회의 시스템의 전 세계 시장규모는 2020년 \$5.77백만달러 대비 2021년 \$6.85백만 달러로 8.1% 성장하였다[1]. 또한 전 세계 약 4억 5천만 명의 이용자가 화상회의 소프트웨어를 이용하여 매일 화상 미팅을 진행하는 것으로 조사되면서 화상회의가 새로운 회의 문화로 자리잡게 되었다[2]. 그러나 화상회의의 사용이 길어지면서 ‘줌 피로(zoom fatigue)’라는 새로운 업무 피로감을 호소하는 사용자가 늘어났다[3]. 일반적으로 화상회의는 카메라를 통해 화면에 자신을 마치 거울을 들여다보는 것과 같은 자기인식을 하게 된다[4]. 그런데 이런 자기인식이 문헌적으로는 의사소통과정에서 부정적인 영향을 미친다고 보고되었을 뿐만 아니라, 생리학적으로 뇌에 부정적인 감정을 초래하는 것으로 밝혀졌다[4].

최근 화상회의 시스템은 사용자가 자신의 모습을 보는 불편함을 극복하기 위해 가상 세계 출현과 함께 대중화된 아바타(avatar) 사용을 적극 검토하고 있다[5]. 전 세계 화상회의 시장의 50%를 차지하는 Zoom은 최근 화상회의에 그림 1(a)과 같은 가상 동물 캐릭터를 아바타로 제공하며 사용자의 동작에 따라 애니메이션시키는 기능을 출시했다[6]. 구글 Meet는 움직이는 3D 아바타 기능을 제공하는데, 그림 1(b)과 같은 가상 캐릭터뿐만 아니라 얼굴형, 피부색, 머리 스타일 등을 조합하는 형태로 아바타를 만들 수 있다[7]. 마이크로소프트의 Teams[8]과 시스코의 Webex[9]는 아바타 기능을 제공하고 있지 않다. 아바타를 제공하는 기존 시스템은 미리 만들어 놓은 동물 또는 사물 캐릭터를 사용하거나 얼굴 요소를 조합하는 방식으로 아바타를 제공하기 때문에 사용자와 아바타 사이의 갭이 커서 화상회의의 현실감이 부족한 단점이 있다. 본 논문에서는 사용자마다 자신을 닮은 아바타를 직접 생성할 수 있고 그 아바타가 사용자의 말하는 모습에 따라 움직이게 하는 기능을 제안한다. 제안하는 기능은 사용자의 얼굴을 아바타로 대체하면서도 사용자와의 현실감을 높일 수 있어 보다 실감나는 화상회의 시스템을 가능케 한다.

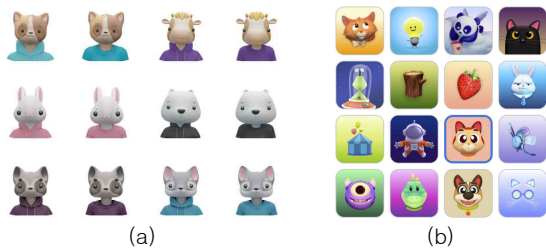


그림 1. 기존 화상회의 시스템의 아바타 예 : (a) Zoom[6]과 (b) 구글 Meet[7]

Fig. 1. Avatar examples of existing video conferencing systems : (a) Zoom[6] and (b) Google Meet[7]

표 1. 화상회의 시스템에 대한 5가지 주요기능 비교 분석
Table 1. Comparative analysis by 5 main functions for video conferencing systems

Main functions	Zoom [6]	Google Meet [7]	Microsoft Teams [8]	Cisco Webex [9]	Proposed
Using an avatar	o	o	x	x	o
Creating an avatar that resembles a user	x	△	x	x	o
Animating an avatar	o	o	x	x	o
Emoji	o	o	o	o	o
Chatbot	x	x	x	x	o

표 1은 기존 화상회의 시스템의 주요 기능에 대해 비교분석하고 있다. 아바타 기능 이외에도 Zoom, Meet, Teams, Webex 모두는 감정을 표현하는 이모티콘은 제공하지만, 아바타의 얼굴과 감정을 표현하고 있지는 않다. 또한 아직 화상회의 일정 관리를 도와주는 챗봇 기능을 제공하는 화상회의 시스템은 없다. 제안하는 화상회의 시스템은 사용자별로 생성한 아바타의 감정 이모티콘을 제공하고 회의채팅 및 커뮤니티에서 활용하는 기능을 제공하며, 회의 관련 질문에 응답하는 챗봇 기능을 구현한다.

II. 관련 연구

본 장은 보다 현실감을 높일 수 있는 아바타를 생성하기 위해 사용된 DualStyleGAN 모델과 아바타 애니메이션을 위해 사용된 Talking head anim(THA) 모델을 살펴본다.

2-1 DualStyleGAN

DualStyleGan은 실제 얼굴 도메인과 초상화 도메인의 이중 스타일을 유연하게 제어하여 자연스러운 초상화 얼굴 합성을 만들어내는 모델이다[10]. DualStyleGAN은 기존 StyleGAN의 내부 스타일 경로에 새로운 외부 스타일 경로를 추가하여 고해상도의 초상화 스타일을 전송한다. 본 논문에서 DualStyleGAN은 화상회의의 사용자와 닮은 아바타를 생성하는 데 사용된다.

그림 2은 DualStyleGAN의 구성도를 보여주고 있다. 얼굴 이미지 I 와 초상화 이미지, 즉 만화 캐릭터 이미지 S 가 주어지면, $G(E(I), E(S), w)$ 에 의해 스타일이 전송된다. 여기서 E 는 인코더(encoder), G 는 생성자(decoder 또는 generator), w 은 내부 스타일 경로와 외부 스타일 경로의 유연한 스타일 조합에 대한 가중치 벡터이다. 색상은 외부 스타일 경로의 매핑 네트워크를 통해 추출된 색상 코드를 도메인별로 색상을 특성화하는 색상 변환(color transform) 블록에 전달되고 미세(fine) 해상도를 위한 생성자에서 w 값을 반영하여 융합된다. 구조는

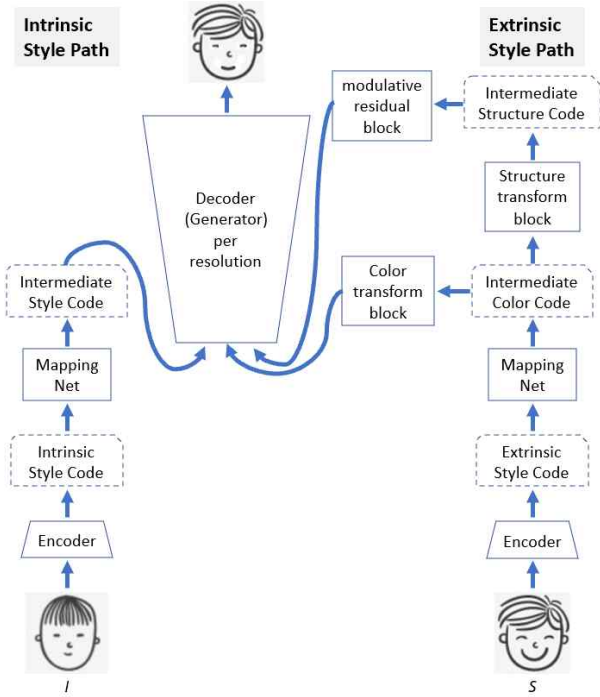


그림 2. DualStyleGAN[10] 구성도
 Fig. 2. Structure of DualStyleGAN[10]

도메인별로 스타일 구조를 특성화하는 구조 변환(structure transform) 블록을 통해 변조 잔차(modulative residual) 블록에 전달되고 거친(coarse) 해상도의 생성자에서 융합된다.

2-2 Talking head anime(THA) 모델

THA 모델은 한 장의 2D 이미지로 애니메이션과 같이 표정을 바꾸고 3D 움직임을 얻을 수 있는 GAN 모델이다[11]. 본 논문에서 THA 모델은 캐릭터 애니메이션 기능과 감정 이모티콘 생성에 사용된다.

THA 모델은 눈썹, 눈, 입, 눈동자 크기 및 위치, 얼굴 회전의 값을 가진 6차원 포즈 벡터를 face morpher와 face rotator 네트워크로 전달해서 캐릭터 이미지를 변경한다. 먼저 face morpher는 눈과 입이 열리는 정도로 표정을 변화시킨다[11]. 식 (1)은 face morpher의 손실함수이다.

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

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

다음으로 face rotator는 얼굴을 회전시킨다. 얼굴 회전은 2D가 아닌 3D 공간에서 이루어지므로 2D 이미지를 3D로 변환시키는 알고리즘이 필요하다. Face rotator는

GANimation[12]과 GridFace[13]의 두 개의 뷰 합성 알고리즘의 결과값을 더하여 얼굴 회전을 계산한다. 식 (2)은 face rotator의 손실 함수이다.

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

여기서 I_p 은 포즈된 이미지이고 I_k 은 GANimation[12]과 GridFace[13]의 네트워크를 통과한 이미지이다. $\Phi_j(I_p, I_k)$ 은 이미지 분류용 알고리즘인 VGG-16을 정의한다[14].

III. 제안하는 아바타 생성 및 애니메이션

본 장에서는 DualStyleGAN과 THA 모델을 이용하여 한국 만화 ‘여신강림’의 캐릭터를 가진 아바타를 생성하고 애니메이션시키며 감정 이모티콘을 생성하는 방법을 제안한다.

3-1 한국 만화 캐릭터를 이용한 아바타 생성

기존 DualStyleGAN 모델이 서구적인 만화 캐릭터들을 대상으로 학습되었다면, 본 논문에서는 ‘여신 강림’이라는 한국 만화 캐릭터를 추가하여 학습한다. ‘여신 강림’은 네이버 웹툰 중 신드롬을 일으켰던 작품으로 네이버, 구글 등의 웹사이트에서 크롤링을 통해 이미지 데이터를 수집하였다. 그림 3은 아바타 생성을 위한 학습에 사용된 ‘여신강림’ 만화 캐릭터 이미지 예시를 보여준다.

아바타 생성을 위한 DualStyleGAN의 학습은 크게 네 개의 과정으로 나뉜다. 첫 번째 과정인 얼굴 디스타일화(destylization)는 예술적 초상화에서 사실적인 얼굴을 복구하는 것으로 고정된 얼굴-만화 캐릭터 쌍을 형성하는 것을 목표로 한다. 그림 4는 얼굴 디스타일화 과정을 보여주는데, 그림 4(a)는 만화 캐릭터 학습 데이터로 그림 1의 S에 해당한다. 그림 4(b)는 S가 인코더와 내부 스타일 경로의 생성자 g에 의해

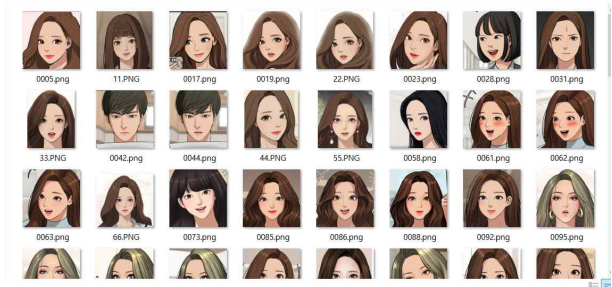


그림 3. 아바타를 생성하기 위한 학습 데이터로 사용된 ‘여신강림’ 만화 캐릭터 이미지 예
 Fig. 3. Examples of ‘True Beauty’ cartoon character images used as learning data to generate an avatar

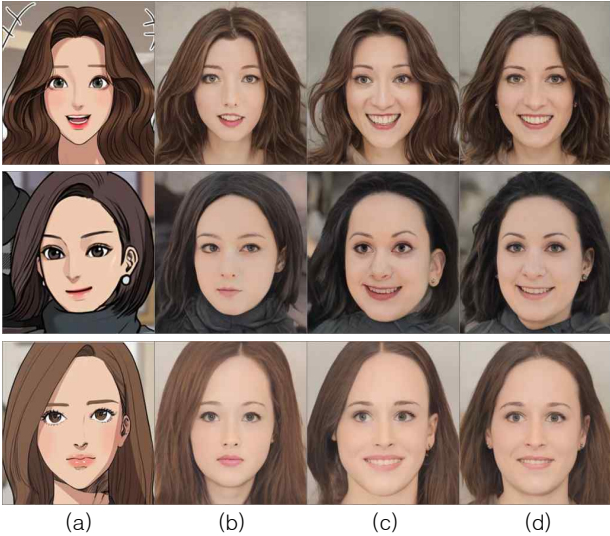


그림 4. 얼굴 디스타일화 결과 : (a) 입력 캐릭터 S , (b) 잠재공간 초기화 얼굴 $g(z_e^+)$, (c) 잠재공간 최적화 얼굴 $g(z_e^+)$, (d) 최종 디스타일화된 얼굴 $g(z_i^+)$

Fig. 4. Results of face destylization : (a) the input character image S , (b) the latent-initialized face $g(z_e^+)$, (c) the latent-optimized face $g(z_e^+)$, and (d) the final destylized face $g(z_i^+)$

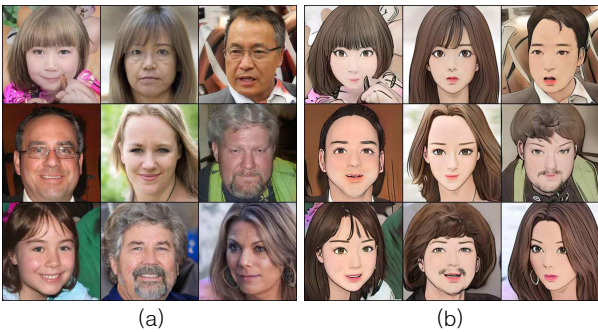


그림 5. '여신 강림'으로 학습된 DualStyleGAN에 의해 생성된 결과 : (a) 입력 얼굴, (b) 생성된 얼굴

Fig. 5. Results generated by DualStyleGAN learned with 'True Beauty' : (a) input real faces and (b) generated faces

StyleGAN의 잠재공간 z^+ 에 초기 매핑된 얼굴을 나타낸다 ($g(E(S)) = g(z_e^+)$). 그림 4(c)는 초기 매핑된 얼굴을 외부 스타일 경로의 생성자로 최적화한 얼굴이며($g(\hat{z}_e^+)$), 그림 4(d)는 최적화한 얼굴에서 비현실적인 얼굴 정보를 추가로 제거한 최종 디스타일화한 얼굴을 보여준다($g(g(\hat{z}_e^+)) = g(z_i^+)$).

아바타 생성을 위한 두 번째 학습 과정은 그림 1의 DualStyleGAN을 통한 학습이다. 내부 스타일 경로는 만화 캐릭터의 디스타일화된 z_i^+ 또는 인코더 E 에 내재된 실제 얼굴 이미지의 잠재공간 z^+ 을 받는다. 외부 스타일 경로는 만화

캐릭터의 초기 얼굴인 z_e^+ 를 사용한다. 그림 5는 '여신강림'에 의해 추가 학습된 DualStyleGAN에 의해 생성된 캐릭터 결과물을 보여준다. 그림 5(a)는 실제 얼굴 이미지이고, 그림 5(b)는 그림 5(a)를 '여신강림' 스타일의 만화 캐릭터 얼굴로 생성된 결과 이미지이다.

아바타 생성을 위한 세 번째 과정은 점진적 미세 조정 과정으로, 목표 도메인의 색상, 얼굴 구조, 얼굴 스타일을 차례대로 기존 도메인에 전송한다. 그림 6은 점진적 미세 조정 과정을 단계별로 보여준다. 그림 6(a)은 입력 얼굴 이미지와 만화 캐릭터(왼쪽 하단)를, 그림 6(b)은 1단계에서 만화캐릭터의 색상을 변환한 결과를, 그림 6(c)은 2단계에서 만화 캐릭터의 구조를 변환한 결과를 보여준다. 그림 6(d)은 3단계에서 최종 미세 조정으로 스타일을 변환한 결과를 보여준다. 아바타 생성을 위한 마지막 과정은 잠재공간에서의 최적화와 샘플링으로, 조정된 모델로 얻은 이미지는 훨씬 스타일 이미지에 가까워진다.

지금까지 설명한 아바타 생성을 위한 네 개의 과정을 포함한 학습은 운영체제는 윈도우 10, CPU는 Intel i5, GPU는 NVIDIA GeForce RTX 2060인 컴퓨터에서 이루어졌다. 얼굴 디스타일화 과정에서는 배치 크기를 1로, 반복을 300으로, 보다 사실적으로 만드는 하이퍼 파라미터인 truncation을 0.7로 설정했다. 점진적 미세 조정 과정에서는 학습률을 0.002로 설정하고, 최적화 함수로 Adam을 사용했다. 미세조정 과정의 1, 2 단계에서 반복은 3000으로, 배치 크기는 4로 설정했고, 손실 함수로 perceptual loss를 사용했다. 3 단계에서는 반복을 1500으로, 배치 크기를 2로 설정했고, 손실함수는 contextual loss를 사용했다. 마지막 최적화와 샘플링 과정에서는 배치 크기를 1로, 반복을 100으로, 색상 변환을 위한 학습률을 0.1로, 구조 변환을 위한 학습률을 0.005로 설정했다. 그림 7은 '여신강림'의 만화 캐릭터를 이용하여 모든 학습이 진행된 후 생성된 사용자와 닮은 만화 캐릭터 이미지를 보여준다. 그림 7의 두 번째 행 가장 왼쪽의 만화 캐릭터 스타일이 첫 번째 행의 다양한 입력 얼굴에 적용되어 생성된 결과물이다.

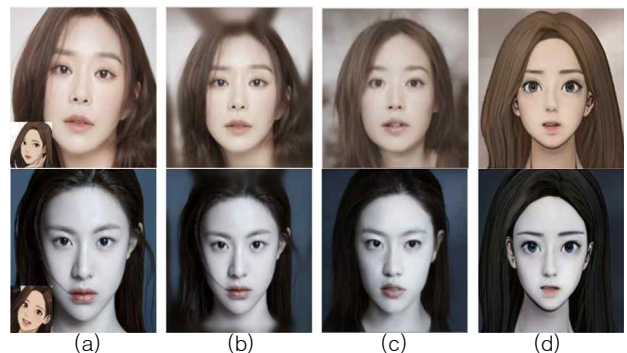


그림 6. 점진적 미세 조정 결과 : (a) 입력 이미지, (b) 1단계, (c) 2단계, (d) 3단계

Fig. 6. Results of progressive fine-tuning : (a) input, (b) after stage I, (c) after stage II, and (d) after stage III



그림 7. 최종 생성된 아바타 결과
Fig. 7. Final results of generated avatars

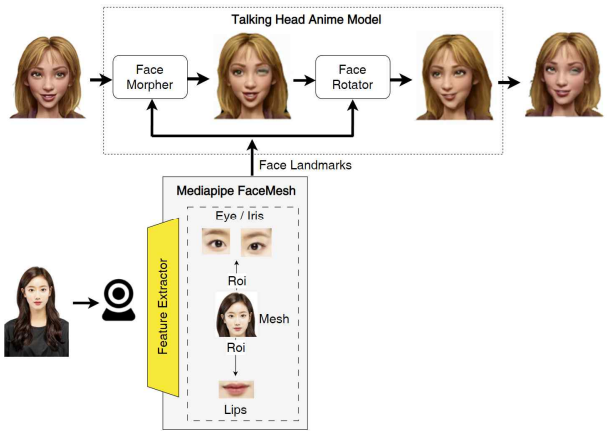


그림 8. 생성된 아바타의 애니메이션 흐름도
Fig. 8. Flowchart for animating the generated avatar

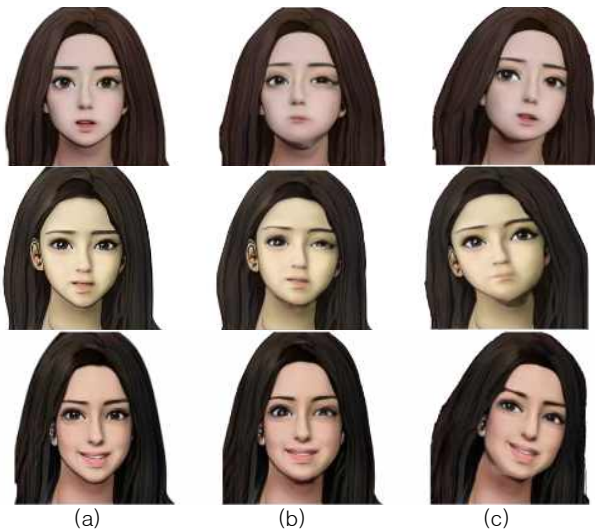


그림 9. THA 모델의 실행 결과 : (a) 2D 이미지, (b) face morpher의 결과, (c) face rotator의 결과
Fig. 9. Results by the Talking head anime model : (a) 2D avatar, (b) face morpher, and (c) face rotator

3-2 아바타 애니메이션

본 절에서는 THA 모델을 이용하여 3-1절에서 생성한 2D 아바타 이미지를 애니메이션시키는 방법을 설명한다. 그림 8은

본 논문에서 구현한 캐릭터 애니메이션을 위한 흐름도를 보여 준다. 먼저 비디오, 오디오 등과 같은 시계열 데이터를 처리하는 기계 학습 파이프라인을 구축하기 위한 프레임워크인 미디어파이프(MediaPipe)를 이용하여[15] 웹캠을 통해 사용자의 움직임을 실시간으로 얻어낸다. 미디어파이프의 얼굴 메쉬(FaceMesh)는 실시간으로 사용자 얼굴의 특징을 추출할 수 있는 솔루션으로, 눈과 눈동자, 눈썹, 코, 입술, 그리고 얼굴 윤곽의 랜드마크를 추출하여 위치를 얻을 수 있다. 이렇게 추출된 얼굴 랜드마크 위치값을 THA 모델에 전달한다. THA 모델은 전달받은 랜드마크 값을 face morpher와 face rotator 네트워크로 전달하여 DualStyleGan을 통해 만들어진 고정된 포즈의 아바타 이미지가 실시간으로 추출된 사용자의 얼굴 랜드마크 값에 따라 실시간으로 움직이는 모습을 보여줄 수 있다. 그림 9는 아바타 이미지(그림 9(a))를 face morpher와 face rotator 네트워크로 실행시킨 결과를 각각 보여준다.

3-3 감정 이모티콘 생성

3-2절과 같이 눈, 코, 입, 귀의 좌표값에 따라 2D 캐릭터를 움직일 수 있는 THA 모델을 이용하여 각 좌표값을 조절함으로써 다양한 표정의 이모티콘을 생성할 수 있다. 우선 256×256의 크기이며 컬러에 알파 채널을 추가한 RGBA 형식으로 배경이 투명한 파일을 준비한다(그림 10). 3-1에서 생성된 2D 아바타는 128×128의 크기로 재조정되어 투명 파일의 중앙에 배치된다.

다음으로 고정된 표정을 가진 2D 아바타의 눈, 코, 입, 귀의 얼굴 좌표값에 ‘기쁨’, ‘슬픔’, ‘화남’, ‘윙크’ 등 4가지 표정을 위한 포즈를 적용한다. THA는 눈썹, 눈, 입, 얼굴 등을 표정이나 움직임에 따라 12개의 양 눈썹 포즈, 12개의 양 눈 포즈, 9개의 입 포즈 등을 제공한다. 각 포즈는 0~1 사이의 실수를 가지며, 1의 값에 가까워질수록 뚜렷한 표정 변화가 나타난다. 본 논문에서는 정확한 표정을 보여주기 위하여 최댓값인 1을 적용한다. 예를 들어, ‘기쁨’의 경우 기쁨의 두 눈과 ‘e’ 모양의 입에 해당하는 포즈값만 1로, 다른 포즈값을 0으로 설정하면 THA 모델은 웃고 있는 눈과 올라간 입 모양을 표현할 수 있다. ‘슬픔’은 곤란함의 두 눈썹과 아래쪽 입꼬리를, ‘화남’은 화남의 두 눈썹과 아래쪽 입꼬리를, ‘윙크’는 기쁨의 왼쪽 눈만 1로 설정하여 캐릭터에 표정을 적용한다. 그림 11은 2D 아바타에 기쁨, 슬픔, 화남, 윙크의 4개의 표정을 적용한 결과를 보여준다.

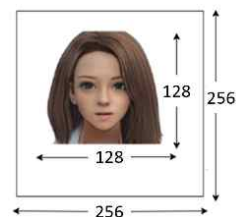


그림 10. 이모티콘 구성도
Fig. 10. Emoticon configuration

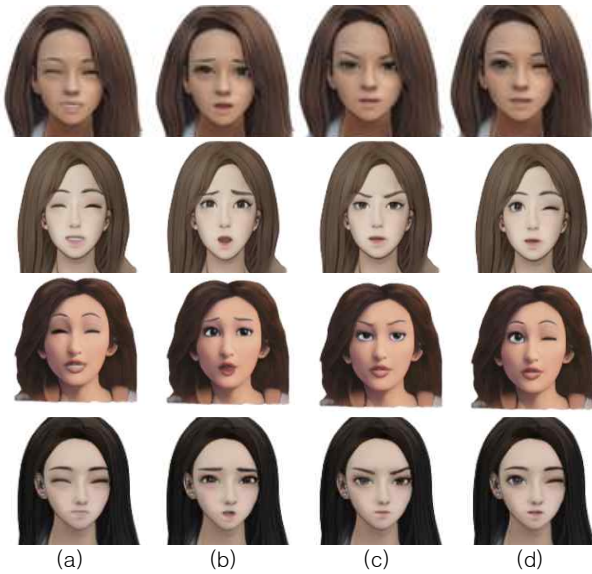


그림 11. 감정 이모티콘 생성 결과 : (a) 기쁨, (b) 슬픔, (c) 화남, (d) 윙크
 Fig. 11. Results of emoticons with emotions : (a) happy, (b) sad, (c) angry, and (d) wink

IV. 화상회의 시스템에 적용

본 장에서는 III 장에서 제안한 사용자별 아바타 생성 및 애니메이션 기능을 적용한 화상회의 시스템을 제안한다. 또한 사용자별 아바타의 감정 이모티콘과 회의 관련 질문에 응답하는 챗봇 기능을 화상회의 시스템에 구현한다.

4-1 제안하는 화상회의 시스템 설계

본 논문에서 제안하는 화상회의 시스템의 아키텍처는 그림 12와 같다. 아바타 생성 모델인 DualStyleGan과 이모티콘, 이미지 애니메이션 생성을 위한 THA 모델은 플라스크(Flask) 서버로 제공하고, 사용자, 일정, 게시글 등은 스프링부트(Spring boot) 서버를 통해 제공한다. 클라이언트는 웹과 모바일 앱으로 제공하여 사용자의 편리성을 보완한다. 웹은 리액트(React)를 사용해 구현하며, 모바일 앱은 안드로이드를 사용해 구현한다. 화상회의는 웹 또는 모바일 앱에서 영상 통화 기능을 쉽게 추가할 수 있는 오픈소스 플랫폼인 오픈바이두(OpenVidu)를 활용해 구현한다[16]. 클라이언트와 스프링부트 서버 사이의 알림은 메시지를 안정적으로 전송할 수 있는 크로스 플랫폼 메시지 솔루션인 Firebase 클라우드 메시징(FCM)으로 구현한다. 데이터베이스는 사용자 정보와 화상회의 정보 등을 MariaDB로, 챗봇 정보를 레디스(redis)로 구현하여 저장한다.

또한 제안하는 화상회의 시스템은 대화식 사용자 인터페이스를 모바일 앱, 웹 등에 설계하고 통합할 수 있는 플랫폼인 구글 다이얼로그플로우(DialogFlow)를 활용해 대화를 구성

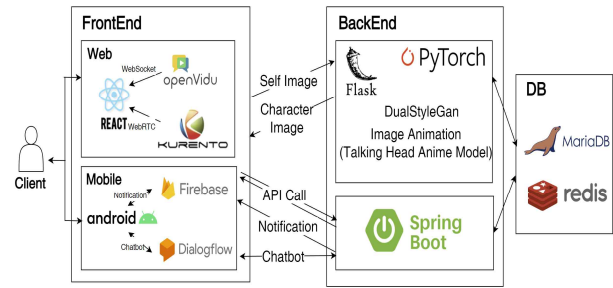


그림 12. 제안하는 화상회의 시스템의 아키텍처
 Fig. 12. Architecture of the proposed video conferencing system

하고 챗봇을 구현한다. 그림 13은 제안하는 챗봇의 인텐트 구조를 보여준다. 화상회의를 위한 챗봇은 일정 조회, 알림, 예약의 3가지 업무를 수행한다.

4-2 제안하는 화상회의 시스템 개발 결과

그림 14는 본 논문에서 제안하는 화상회의 시스템 개발 결과를 보여준다. 화상회의 사용자는 아바타를 생성하기 위해 얼굴을 포함하는 사진을 업로드하거나 카메라 사진 촬영 방식으로 직접 입력받아올 수 있다(그림 14(a)). 생성될 아바타의 이름과 설명을 입력한 후 원하는 만화 캐릭터 종류를 선택한다(그림 14(b)). 선택할 수 있는 만화 캐릭터의 종류는 만화, 캐리커처, 애니메이션, 아케인, 픽사, 여신강림으로, 그림 14(c)는 만화에 포함된 70개의 캐릭터와 여신강림에 포함된 16개의 캐릭터를 보여준다. 그림 14(d)는 여신강림 13번 캐릭터를 선택했을 때, 그림 14(a)의 입력 영상으로 생성된 아바타와 이 아바타의 기쁨, 슬픔, 화남, 윙크의 표정을 적용한 이모티콘을 보여준다.

사용자는 화상회의를 사용하기 위해 캘린더에서 원하는 날짜와 시간에 회의실을 예약한다. 회의실에 입장하기 전에 미리 생성해 놓은 아바타 중 하나를 선택한다(그림 14(e)). 회의장에 참여하면 선택된 아바타가 사용자를 대신해서 화면에 보이며 회의 중 카메라에 비친 사용자의 움직임에 따라 아바타가 움직인다.(그림 14(f)). 회의 중 음소거, 비디오 중지, 회의 나가기와 같은 기능이 제공된다. 회의실의 오른쪽엔 채팅방이 있고 아바타와 함께 생성된 감정 이모티콘을 사용해 대화할 수 있다. 또한 회의실에 참여하는 사용자 목록을 확인할 수 있으며(그림 14(g)), 일정 제목, 일정 타입, 시간, 내용, 알람시간 등의 정보를 입력하여 회의 일정을 추가하고 개설할 수 있다(그림 14(h)).

제안하는 화상회의 시스템은 챗봇 기술을 접목해 사용자의 전체적인 일정을 관리한다(그림 14(i)). 챗봇을 이용하여 팀 이름, 회의 일정 등에 대한 안내를 통해 회의를 예약하고 알람을 설정할 수 있다. 또한 팀 이름과 회의 일정을 이용하여 예약된 회의를 조회할 수도 있다.

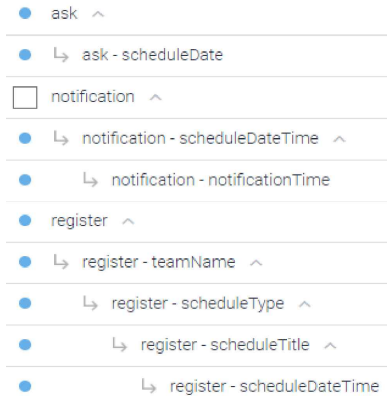
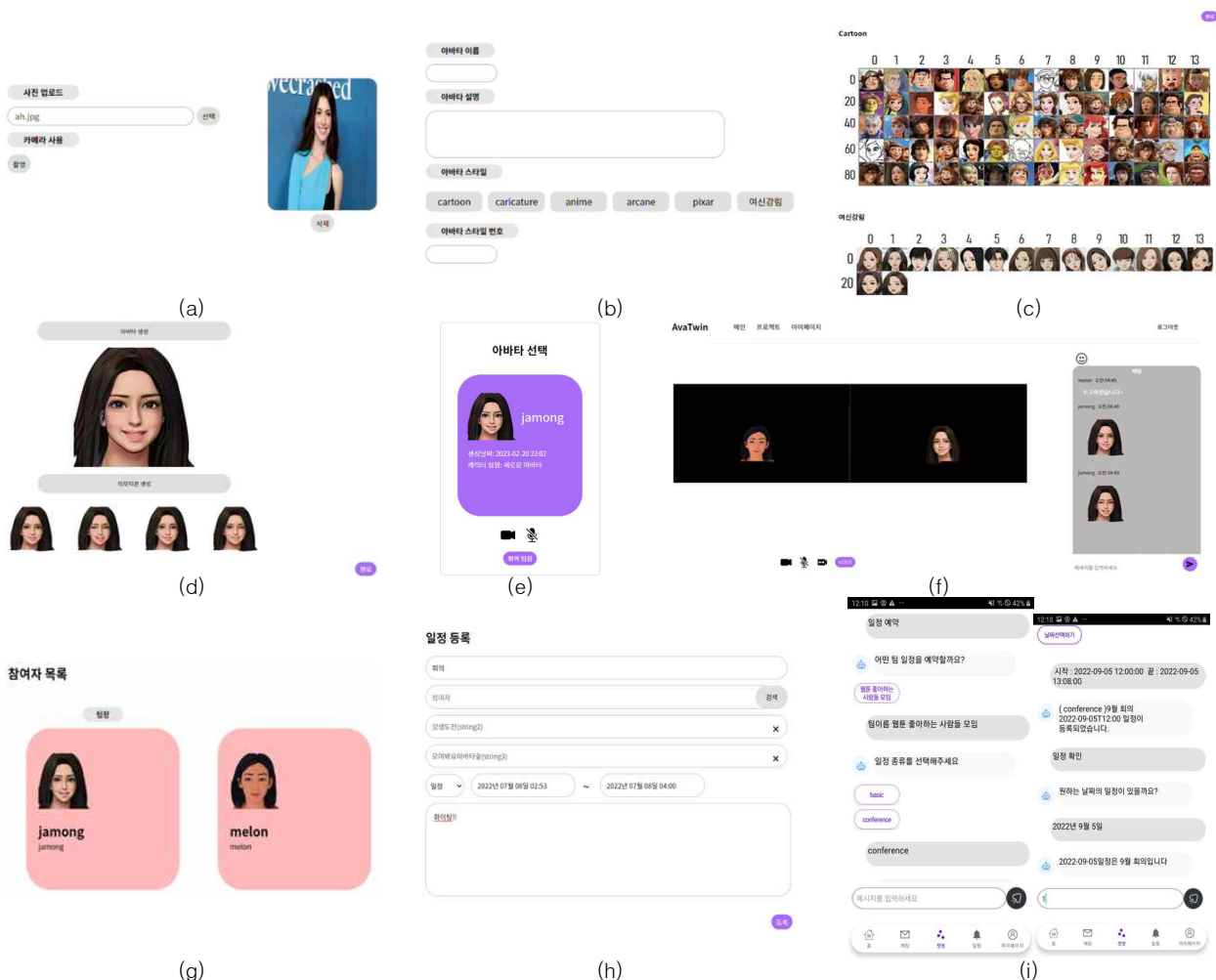


그림 13. 제안하는 챗봇의 인텐트 구조
Fig. 13. Structure of intent in the proposed chatbot

V. 결 론

본 논문에서 제안하는 화상회의 시스템은 DualStyleGAN과 THA 모델을 이용하여 자신과 닮은 한국형 아바타를 생성하였고, 사용자의 움직임에 따라 아바타를 애니메이션시켰으며, 그 아바타가 감정을 표현하게끔 이모티콘을 생성하여 화상회의에 활용될 수 있게 구현되었다. 기존 화상회의 시스템은 시스템이 미리 만들어 제공하는 동물 캐릭터나 사람 얼굴의 요소들을 조합하는 방식으로만 아바타를 사용할 수 있어 아바타와 사용자 사이의 갭이 크다는 단점을 가지고 있다. 반면, 제안하는 시스템은 자신의 얼굴과 만화캐릭터를 이용하여 직접 자신과 닮은 아바타를 생성할 수 있다. 이는 사용자의 얼굴을 아바타로 가리면서도 아바타와 실제 사용자 사이의



*These figures are screen captures of the developed system, and this system is made for Korean users.
그림 14. 제안하는 화상회의 시스템 화면 : (a) 입력 얼굴(사진 또는 카메라), (b) 아바타 정보 입력(이름, 설명, 종류), (c) 만화 캐릭터들, (d) 아바타 생성 결과(아바타와 4개 감정 이모티콘), (e) 화상회의를 위한 아바타 선택, (f) 화상회의 화면(아바타, 채팅, 이모티콘 등), (g) 회의 참여자 목록, (h) 회의 일정 등록, (i) 회의 일정 등록 및 검색을 도와주는 챗봇
Fig. 14. Screenshots of the proposed video conferencing system : (a) input face(photo or camera), (b) avatar information (name, description and types), (c) cartoon characters, (d) avatar generation result (avatar and 4 emotional emoticons), (e) selection of an avatar for video conferencing, (f) screen of video conference system (avatars, chats, emoticons, etc.), (g) list of meeting participants, (h) registration of meeting schedule and (i) chatbot that helps register and search meeting schedules

캡을 줄일 수 있어 회의 참가자들에게 보다 몰입감을 주고 화상회의의 실제감을 높이는 환경이 조성되도록 도움을 준다. 또한 이모티콘을 활용하여 감정을 이미지화하여 전달해줌으로써 말로만 서술하는 것보다 자기 생각이나 감정을 더 효과적으로 전달할 수 있을 것이다.

본 논문은 최근 가장 많이 연구되고 있는 이미지 변환/생성 기술 중 하나인 얼굴 생성(face generation) 기술을 웹 화상회의 응용 분야에 적용하여 상용화의 가능성을 보여주었다. 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 활용될 수 있을 것이다. 향후 사용자가 선호하는 만화 캐릭터를 추가로 학습하여 제공할 것이며, 나를 닮은 3D 아바타 생성, 음성인식에 기반을 둔 애니메이션 등의 기능들을 추가할 계획이다.

참고문헌

[1] Precedence Research. ICT – Video Conferencing Market [Internet]. Available: <https://www.precedenceresearch.com/video-conferencing-market>

[2] RSupport. Video Conferencing Market Trends [Internet]. Available: https://files.rsupport.com/kr/remotemeeting/documents/whitpapers/remotemeeting_video_conferencing_trend.pdf

[3] E. J. Lee, “Video-conference Fatigue of College Students,” *The Journal of the Convergence on Culture Technology*, Vol. 8, No.3, pp. 589-594, 2022. <http://dx.doi.org/10.17703/JCCT.2022.8.3.589>

[4] R. Riedl, “On the Stress Potential of Video Conferencing: Definition and Root Causes of Zoom Fatigue,” *Electronic Markets*, Vol. 32, pp. 153-177, March, 2022. <https://doi.org/10.1007/s12525-021-00501-3>

[5] P. Panda, M. J. Nicholas, M. Gonzalez-Franco, K. Inkpen, E. Ofek, R. Cutler, ... and J. Lanier, “AllTogether: Effect of Avatars in Mixed-Modality Conferencing Environments,” in *Proceeding of 2022 Symposium on HCI for Work*, Durham, NH, USA, pp. 1-10, June 2022. <https://doi.org/10.1145/3533406.3539658>

[6] Zoom [Internet]. Available: <https://www.zoom.us>

[7] Google Meet [Internet]. Available: <https://meet.google.com>

[8] Microsoft Teams [Internet]. Available: <https://www.microsoft.com/microsoft-teams>

[9] Cisco Webex [Internet]. Available: <https://www.webex.com>

[10] S. Yang, L. Jiang, Z. Liu, and C. C. Loy, “Pastiche Master: Exemplar-based High-resolution Portrait Style Transfer,” in *Proceeding of IEEE/CVF Conference on CVPR*, New

Orleans, Louisiana, USA, pp. 7693-7702, June 2022. <https://doi.org/10.48550/arXiv.2203.13248>

[11] P. Khungurn. Talking Head Anime from a Single Image 2: More Expressive [Internet]. Available: <https://github.com/pkhungurn/talking-head-anime-2-demo>, 2021.

[12] 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*, Munich, Germany, pp. 835-851, September 2018. https://doi.org/10.1007/978-3-030-01249-6_50

[13] E. Zhou, Z. Cao, and J. Sun, “GridFace: Face Rectification via Learning Local Homography Transformations,” in *Proceedings of the Conference on ECCV*, Munich, Germany, pp. 3-20, September 2018. <https://doi.org/10.48550/arXiv.1808.06210>

[14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition,” *arXiv Preprint arXiv:1409.1556*, 2014. <https://doi.org/10.48550/arXiv.1409.1556>

[15] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, ... and M. Grundmann, “MediaPipe: A Framework for Building Perception Pipelines,” in *Proceedings of Third Workshop on Computer Vision for AR/VR at IEEE CVPR*, Long Beach, June 2019. <https://doi.org/10.48550/arXiv.1906.08172>

[16] OpenVidu [Internet]. Available: <https://openvidu.io/>



안효원(Hyo-Won Ahn)

2023년 : 덕성여자대학교 컴퓨터공학과 (공학사)

2023년~현 재: 고려대학교 컴퓨터학과 학부과정
※ 관심분야 : 영상처리, 컴퓨터비전, 데이터 마이닝



박지혜(Ji-Hye Park)

2023년 : 덕성여자대학교 컴퓨터공학과 (공학사)

※ 관심분야 : 모바일 앱, 서버



김지윤 (Ji-Yun Kim)

2023년 : 덕성여자대학교 컴퓨터공학과(공학사)

※ 관심분야 : 웹, 모바일 앱, 데이터베이스



문지연 (Ji-Eon Moon)

2023년 : 덕성여자대학교 컴퓨터공학과(공학사)

※ 관심분야 : 웹, 클라우드, 데이터베이스



이경미 (Kyoung-Mi Lee)

1993년 : 덕성여자대학교 전산학과(이학사)

1996년 : 연세대학교 전산과학과(이학석사)

2001년 : 아이오와 대학교 전산학과(전산학박사-지능형 멀티미디어)

2003년~현 재: 덕성여자대학교 컴퓨터공학과 교수

※ 관심분야 : 영상처리, 컴퓨터비전, 패턴인식