

일반논문 (Regular Paper)

방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)

<https://doi.org/10.5909/JBE.2022.27.3.369>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

피부 미세요소 조절을 통한 얼굴 영상 합성

김 유 진^{a)}, 박 인 규^{a)†}

Facial Image Synthesis by Controlling Skin Microelements

Yujin Kim^{a)} and In Kyu Park^{a)†}

요 약

최근 딥러닝 기반의 얼굴 합성 연구는 전체적인 스타일이나 헤어, 안경, 화장과 같은 요소를 포함하는 매우 사실적인 얼굴을 생성하는 결과를 보인다. 그러나 피부의 미세 구조와 같은 매우 세부적인 수준의 얼굴은 생성하지 못한다. 본 논문에서는 이러한 한계점을 극복하고자 한 장의 얼굴 라벨 영상으로부터 피부 미세 요소의 종류와 강도 조절을 통해 더욱 사실적인 얼굴 영상을 합성하는 기법을 제안한다. 제안하는 기법은 Image-to-Image Translation 방법인 Pix2PixHD를 이용해 얼굴 영역과 피부 요소인 주름, 모공, 홍조가 표시된 라벨 영상을 변환하여 미세 요소가 추가된 얼굴 영상을 생성한다. 피부 요소 영역을 조절한 라벨 영상을 다양하게 생성함으로써 이에 대응하는 미세한 피부 요소가 반영된 다양한 사실적인 얼굴 영상을 생성할 수 있음을 실험을 통해 보인다.

Abstract

Recent deep learning-based face synthesis research shows the result of generating a realistic face including overall style or elements such as hair, glasses, and makeup. However, previous methods cannot create a face at a very detailed level, such as the microstructure of the skin. In this paper, to overcome this limitation, we propose a technique for synthesizing a more realistic facial image from a single face label image by controlling the types and intensity of skin microelements. The proposed technique uses Pix2PixHD, an Image-to-Image Translation method, to convert a label image showing the facial region and skin elements such as wrinkles, pores, and redness to create a facial image with added microelements. Experimental results show that it is possible to create various realistic face images reflecting fine skin elements corresponding to this by generating various label images with adjusted skin element regions.

Keyword : Face synthesis, Image-to-Image Translation, Skin elements

a) 인하대학교 전기컴퓨터공학과(Inha University, Department of Electrical & Computer Engineering)

† Corresponding Author : 박인규(In Kyu Park)

E-mail: pik@inha.ac.kr

Tel: +82-32-860-9190

ORCID: <https://orcid.org/0000-0003-4774-7841>

※이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2020-0-01389, 인공지능융합연구센터지원(인하대학교))과 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019R1A2C1006706).

※This work was partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-01389, Artificial Intelligence Convergence Research Center (Inha University)) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1A2C1006706).

· Manuscript received March 30, 2022; Revised April 7, 2022; Accepted April 7, 2022.

I. 서 론

얼굴 영상 합성은 주어진 의미론적 영역을 통해 사실적인 얼굴 영상을 생성하는 분야이다. 얼굴 초해상화, 정면화, 스타일 변환과 같은 응용 연구가 진행되고 있다. 딥러닝 기반 방법으로 GANs(Generative Adversarial Networks)^[2]이 제안된 이후 타깃 데이터 분포에 맞게 학습할 수 있게 되었다. GANs은 생성자와 판별자를 경쟁적으로 학습시켜 새로운 영상을 합성하는 방법이다. 이 후 GANs에 조건을 추가하여 영상을 생성하는 방식인 CGAN(Conditional GAN)^[3]이 제안되었다. 얼굴 영상을 대상으로 한 연구는 주로 나이, 성별, 머리 색과 같은 얼굴 내 의미 있는 특징을 변경하여 영상을 생성한다.

CGAN 기반의 Image-to-Image Translation은 한 도메인에서 영상을 가져와 다른 도메인의 영상 스타일을 갖도록 변환하는 분야이다. 얼굴 영상을 대상으로 하는 스타일 변환은 전체적인 스타일이나 헤어, 안경, 화장과 같은 얼굴의 일부 요소를 변환시키는 연구^[4,5,6]들이 존재한다. 그러나 기존 연구들은 피부와 같이 미세한 수준의 변환에는 미치지 못한다는 단점이 있다. 주름을 추가하여 나이를 변화시키는 연구^[7]가 존재하나 사실적인 고해상도 영상을 합성하지 못하며 피부의 세부적인 요소를 표현하지 못하였다. 또한 주름, 모공, 홍조와 같은 피부 요소를 합성하는 연구는 전무하다.

본 논문에서는 CGAN 기반의 신경망 모델을 이용해 피부 요소 조절을 통한 다양한 얼굴 영상 생성 기법을 제안한다. 얼굴 영상 합성을 위해 고해상도 영상 합성 모델인 Pix2PixHD^[8]를 사용하였고, 피부 요소를 추출하기 위해 피에스아이플러스(주)의 Mark-Vu 피부 분석기^[1]를 활용해 데이터를 취득하였다. 피부 요소는 주름, 모공, 홍조를 대상으로 하며 고해상도 얼굴 RGB 영상과 그에 따라 추출된 피부 요소 영상을 취득하였다. 입력이 되는 라벨 영상을 만들기 위해 얼굴의 눈썹, 코, 입, 피부 위치를 찾아 얼굴의 구조를 유지하게 하였다. 존재하는 피부 요소의 위치를 기준으로 위, 아래, 왼쪽, 오른쪽 방향으로 피부 요소를 추가하는 방법과 존재하는 피부 요소를 피부 라벨로 변경하여 피부 요소를 제거하는 방법을 이용해 다양한 라벨 영상을 생성할 수 있다. 이를 통해 각 피부 요소 증가 및 감소 비율을 조절

하여 다양한 피부를 표현할 수 있다. 본 논문의 기여를 요약하면 다음과 같다.

- Mark-Vu 피부 분석기를 통한 고해상도 얼굴 영상과 피부 요소 데이터 취득
- 기존 얼굴 영상 합성 기법의 미세 수준 합성이 부족하다는 단점을 극복하고자 피부 미세요소 조절을 통한 얼굴 합성 기법 제안
- 피부 요소 영역을 조절한 라벨 영상을 다양하게 생성함으로써 이에 대응하는 미세한 피부 요소가 반영된 다양한 사실적인 얼굴 영상을 생성

본 논문의 전체 구성은 다음과 같다. 2장에서는 Image-to-Image Translation과 얼굴 합성에 대한 기존 연구를 소개하며, 3장에서는 본 논문에서 제안하는 피부 미세요소 조절을 통한 얼굴 합성 기법에 대해 설명한다. 4장에서는 제안한 방법의 우수성을 검증하기 위해 정량적, 정성적으로 평가하고 다른 라벨 종류를 이용해 훈련된 모델 결과를 비교한다. 추가적으로 피부 요소 조절 방법을 통해 다양한 얼굴 영상 합성 결과를 보여준다. 마지막으로, 5장에서는 본 논문의 결론을 맺는다.

II. 관련 연구

Image-To-Image Translation은 입력 영상을 타깃이 되는 스타일로 변환하여 영상을 합성하는 분야이다. 대표적으로 Pix2Pix^[9]는 skip connection을 이용한 layer로 이루어진 생성자와 PatchGAN^[10] 기반의 판별자를 이용해 다양한 스타일을 가진 영상 변환을 가능하게 한다. 그러나 256×256 해상도와 같이 저해상도 영상 합성에서만 좋은 결과를 보인다는 단점이 있다. CycleGAN^[11]은 Pix2Pix와 달리 짹이 없는 두 스타일의 영상만으로 훈련 가능한 모델이다. 하나의 도메인에서 다른 도메인으로 변환하는 GANs과 이에 역변환하는 GANs 구조로 이루어져 있다. 적대적 손실 함수와 추가적으로 L1 cycle consistency loss를 사용하여 스타일 변환에 도움을 준다. 훈련 데이터셋에 대한 제약이 줄었으며 양방향 영상 합성이 가능하다는 장점이 있다.

얼굴 영상 합성 연구로 GANs^[2] 방법이 제안된 후 여러 연구들이 제안되었다. 그 중에서도 StyleGAN^[6] 모델이 독보적으로 사실적인 얼굴 영상을 생성한다. StyleGAN은 PGGAN^[12]을 기반으로 하여 스타일 변환을 적용해 각기 다른 스타일을 여러 크기로 변경해 학습하는 방식이다. 이 방법을 통해 서로 다른 스타일을 잘 분리시켜 입력으로 들어간 얼굴 영상을 타깃 얼굴 영상의 스타일에 맞게 변환한다. 이후 제안된 모델로 기존 연구의 생성 결과에서 물방울 형태의 blob이 관찰되는 것을 개선한 StyleGAN2^[13]과 얼굴 영상 데이터가 부족하다는 점을 해결하기 위해 데이터 증강 기법을 소개한 StyleGAN2-Ada^[14]가 있다.

얼굴 합성 분야에서 머리카락의 색과 스타일, 나이, 표정 등과 같은 얼굴 요소를 편집하는 많은 연구가 제안되었다. IcGANs^[15]은 CGAN^[3] 기반 모델로, 얼굴을 대상으로 하는 조건부 합성 초기 연구이다. 네트워크는 크게 인코더와 CGAN으로 이루어져 있으며 두 개의 인코더를 통해 실제 영상을 latent representation과 속성 정보로 각각 변환한다. MNIST^[16]와 CelebA^[17] 데이터셋을 통한 실험 결과로 얼굴의 속성을 편집하는 연구의 가능성을 보여주었다. 이후 제안된 한 개의 특징만을 학습하여 하나의 도메인으로만 변환이 가능한 기존 연구와는 달리 StarGAN^[4]은 하나의 신경망을 통해 여러 도메인으로 변환이 가능하다. 판별자는 생성자가 만든 영상의 진위 유무를 판단하고 도메인 종류를

분류한다. 이로 인해 각 도메인에서 모델을 훈련하지 않고 하나의 모델로 문제를 해결할 수 있어 연산량을 줄일 수 있다. 더 나아가 타깃이 되는 얼굴 속성만 편집하는 AttGAN^[18]이 제안되었다. 판별자가 속성을 분류한다는 점에서 StarGAN과 비슷하지만 AttGAN은 latent representation과 속성 간의 관계를 모델링한다는 점에서 다르다. StarGAN과 AttGAN은 각 속성에 맞게 얼굴 영상이 편집되지만 잡음이 생기거나 흐릿해져 사실적인 합성을 하기 부족하다. 또한 연구에서 선글라스, 수염, 앞머리 스타일과 같은 스타일 변환이 가장 미세한 변환으로, 피부 수준의 변환 까지는 구현하지 못했다. 그리고 사전 훈련된 GAN 모델을 이용해 의미론적 얼굴 편집을 하는 InterFaceGAN^[5]이 제안되었다. 영상 편집을 위해 GAN inversion 방법과 인코더 기반 모델을 사용하는데 latent code를 변경하여 얼굴의 속성을 조작한다.

III. 피부 미세요소 조절을 통한 얼굴 영상 합성 기법

1. 학습 데이터 취득

제안하는 기법은 얼굴 영상과 그에 따른 피부 요소 추출

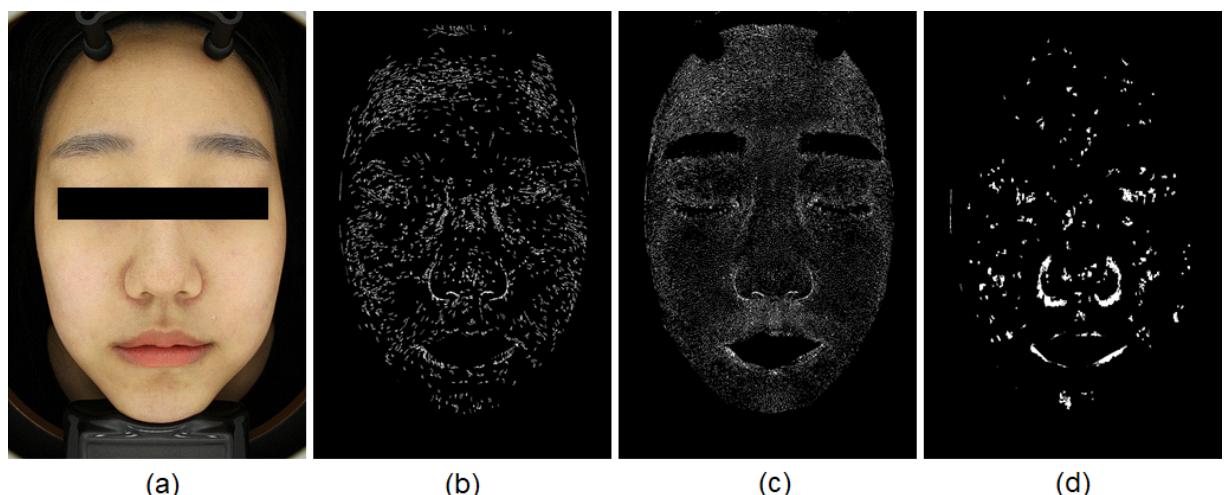


그림 1. 취득한 데이터 예시. (a) 얼굴 정면 영상, (b) 주름 영상, (c) 모공 영상, (d) 홍조 영상
Fig. 1. Examples of data. (a) Face front image, (b) Wrinkle image, (c) Pore image, (d) Redness image

영상이 필요하다. 하지만 얼굴 영상에서 피부 요소를 추출하기 위해서는 피부 요소가 픽셀 단위이고 전문적인 지식을 필요하기 때문에 많은 시간과 비용이 소요된다. 따라서 Mark-Vu 피부 분석기를 이용하여 데이터를 취득한다. 카메라는 CANON 100D DSLR 모델이고 조명에는 일반광, 자외선광, 광택광, 편광이 있으며 본 연구에는 일반광과 편광 영상으로부터 추출된 피부 요소를 이용한다. 다음 과정을 통해 영상 취득이 이루어진다. 먼저, 얼굴 전체를 보이기 위해서 머리카락을 고정하고 피부 분석기 내부의 이마와 턱 받침대에 얼굴을 고정시킨다. 그리고 외부 빛을 차단한 채 자동으로 조명이 바뀌며 4개의 영상이 취득된다. 일반광 영상으로부터 주름과 모공이 추출되며 편광 영상으로부터 홍조가 추출된다. 그림 1은 취득한 데이터 예시를 보여주며 얼굴 정면 영상에 따라 추출된 피부 요소 영상을 나타낸다. 영상 합성에는 일반광에서 취득한 RGB 영상과 주름, 모공, 홍조의 위치와 크기가 표시된 마스크 영상이 이용된다.

남자 91명과 여자 93명을 대상으로 2904×4354 해상도의 얼굴 정면 영상을 취득하였다. 그림 2를 통해 데이터의 연령별 분포 그래프를 나타냈다. 20세부터 57세까지 분포하며 20대 데이터가 가장 많다. 참여자 모두 맨 얼굴로 촬영하여 피부 요소 분석에 용이하게 하였다. 피부 분석기 내 알고리즘으로 산출된 통계로는 피부 유형에서 지성이 가장

많고 다음으로 복합성, 건성, 중성 순이다.

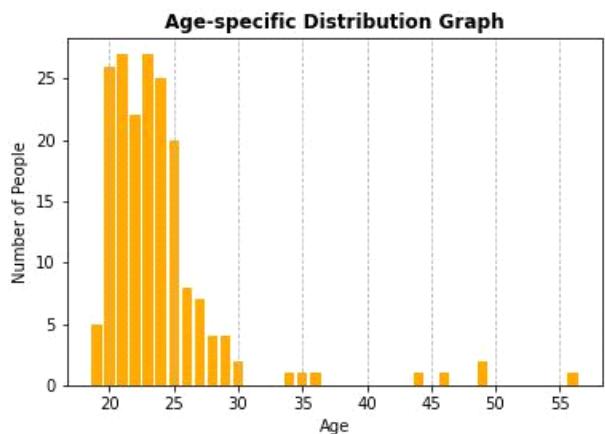


그림 2. 연령별 분포 그래프
Fig. 2. Age-specific distribution graph

2. 제안하는 얼굴 합성 기법

본 절에서는 얼굴 RGB 영상과 미세요소 영상이 입력으로 들어왔을 때 미세요소 조절 기법을 이용해 얼굴 영상을 합성하는 과정에 대해 자세히 서술한다. 제안하는 기법은 Face Parsing, 미세요소 조절, 얼굴 합성의 세 단계로 나눠진다. 제안하는 얼굴 합성 기법 과정은 그림 3에 도시하였

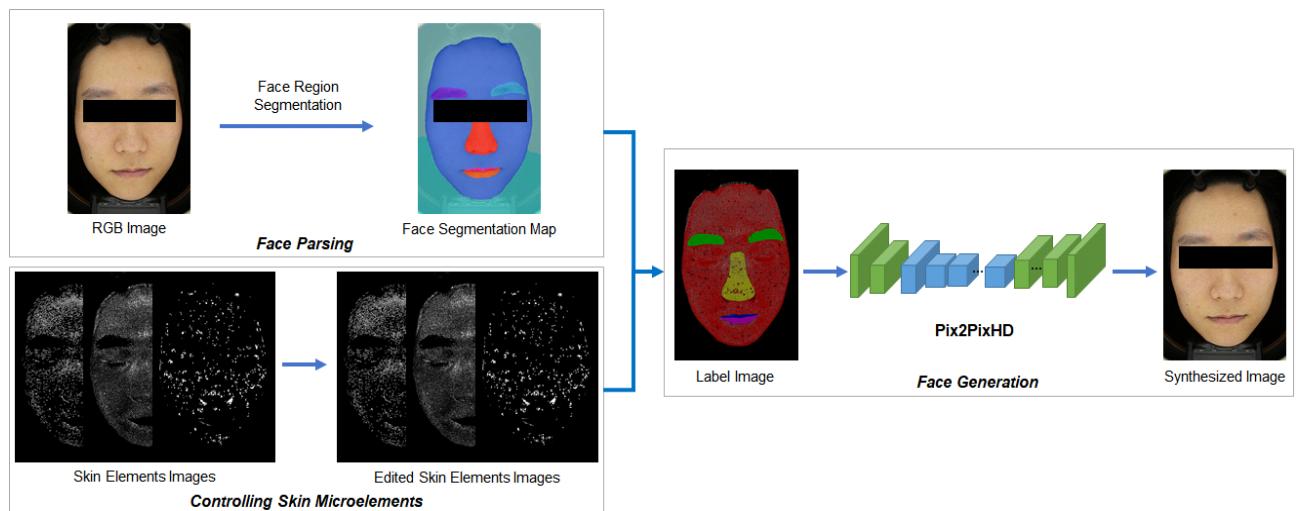


그림 3. 제안하는 피부 미세요소 조절 기법을 적용한 얼굴 합성의 전체 파이프라인
Fig. 3. Overview of the proposed facial synthesis using face microelements control technique

다.

본 논문에서는 미세한 피부를 표현하는 얼굴 영상을 합성하기 위해서 고해상도 합성 모델인 Pix2PixHD^[8]를 이용한다. 훈련을 위해서는 얼굴 RGB 영상과 라벨 영상이 필요하다. 라벨 영상 생성을 위해 먼저 CelebAMask-HQ 데이터셋^[19]으로 훈련된 BiSeNet^[20] 기반 Face parsing 모델을 이용해 얼굴 내 피부, 눈썹, 코, 윗입술, 아랫입술 영역을 추출한다. 이는 얼굴 형태를 유지하기 위함으로 사실적인 피부 요소 합성에도 도움을 준다. 또한, 피부 분석기를 통해 추출된 피부 요소 영상을 이용하여 라벨 개수가 10개인 라벨 영상을 생성한다. 최종적으로 생성되는 라벨의 종류는 배경, 피부, 눈썹, 코, 윗입술, 아랫입술, 주름, 모공, 홍조, 주름 및 모공이다. 피부 분석기로 추득된 피부 요소 추출 영상에 대해 Face Parsing 방법으로부터 얻어진 피부, 눈썹, 입술의 위치를 이용해 전처리 과정을 거친다. 눈썹 영역은 텔로 이루어져 있으며 입술을 홍조라고 판단하는 것을 막기 위해 눈썹과 입술 영역에 피부 요소를 제거하였다. 또한, 주름과 모공이 겹치는 영역이 많기 때문에 한 픽셀에 동시에 존재할 경우를 고려하여 라벨을 추가하였다.

피부 요소 영역을 조절한 라벨 영상을 다양하게 생성하는 라벨 편집 기법을 제안한다. 라벨 편집은 얼굴 내 영역은 고정한 채 피부 요소 라벨만 수정하며, 피부 요소 증가와 피부 요소 감소로 나눌 수 있다. 먼저, 피부 요소 증가 시 주름, 모공, 홍조와 같은 피부 요소는 일반적으로 기준 위치를 기준으로 변화하는 양상을 보인다. 따라서 이 점을 이용하여 피부 요소가 존재하는 픽셀 기준으로 위, 아래, 왼쪽, 오른쪽 픽셀에 라벨을 추가한다. 랜덤으로 확률을 주어 피부 요소가 존재하지 않을 시 0, 존재할 시 1 값을 갖게 한다. 피부 요소 감소 시에는 기존 피부 요소가 존재했던 픽셀 값을 랜덤으로 확률을 주어 0 또는 1 값을 갖게 하였다. 랜덤 확률을 설정할 시 여러 단계로 피부를 합성할 수 있으며 본 논문에서는 총 3단계로 실험하였다. 제안한 기법을 사용할 시에는 한 장의 라벨 영상으로부터 각 피부 요소별로 단계마다 다른 라벨 영상을 생성할 수 있다. 이에 대응하여 미세 피부 요소가 포함된 다양한 사실적인 얼굴 영상을 합성할 수 있다. 또한, 기존 피부 요소 위치를 사용하여 라벨 영상을 편집하는 것은 피부 내 무작위로 추가하는 것보다

더 사실적인 얼굴 영상이 생성되며 실제 피부 변화 양상을 표현할 수도 있다.

Pix2PixHD 모델은 Pix2Pix^[9]을 기반으로 하여 학습 과정이 안정화하고 고해상도 영상 생성을 가능하게 하였다. 본 논문에서는 Pix2PixHD를 얼굴 합성 모델로 이용하는데, 훈련 시에는 취득한 RGB 영상과 라벨 영상이 짹이 되어 입력으로 들어간다. 라벨 영상으로부터 얼굴 RGB 영상을 합성한다. 네트워크 구조는 생성자와 판별자로 구성되어 있으며, 글로벌 생성자와 판별자 2개를 사용하였다. 이를 통해 원본 해상도와 원본 해상도의 절반 크기에서 손실 함수를 계산하여 영상 안에서 전체적으로 일관성 있으면서 작은 디테일을 생성하는 방식을 이용하였다. 손실 함수는 기존 GAN loss와 feature matching loss를 합친 형태이다. 훈련에 instance map은 사용하지 않았으며 본 논문에서 제안하는 라벨 영상만을 이용하였다.

IV. 실험 결과

1. 실험 환경

본 논문에서는 Mark-Vu 피부 분석기를 이용해 취득한 영상을 훈련과 테스트에 사용하였다. 총 184명 참가자 중 훈련 데이터로 164명에 대한 영상을 사용하였으며 테스트를 위해 20명에 대한 영상을 사용하였다. 사용된 영상 종류로는 RGB 영상과 주름, 모공 및 홍조 추출 영상이 있다. 원본 영상의 해상도는 2904×4354 로 크기를 조절할 시 피부 정보가 손실되므로 512×512 해상도를 가진 패치로 잘라 사용하였다. 훈련 데이터는 총 7,544장이며 테스트 데이터는 920장이다.

본 연구에서는 얼굴 생성 모델로 Pix2PixHD^[8]를 사용하는데, 저자가 공식적으로 공개한 코드를 바탕으로 훈련이 진행되었다. 학습 파라미터로 옵티마이저(Optimizer)는 아담(Adam)을 사용하였다. 학습률(Learning rate)은 0.0002로, batch size는 16, epoch는 200로 설정하여 Intel i9-10900KF 3.7GHz CPU와 RTX A6000 GPU를 장착한 컴퓨터 환경에서 훈련하였다.

2. 정성 및 정량 평가

본 절에서는 제안하는 기법과 얼굴 영상 라벨 생성 시 Face Parsing 방법을 사용하지 않을 때의 정성적 및 정량적 평가 및 비교 분석을 작성하였다. 또한 한 장의 라벨 영상으로 피부 요소가 있을 확률을 조절하여 다양한 라벨 영상을 생성함으로써 얼굴 영상을 생성할 수 있음을 보였다.

제안하는 기법을 평가하기 위해 라벨 영상을 입력으로 받아 얼굴 합성 결과를 실제 영상과 비교하여 그림 4에서 나타냈다. 얼굴 영역별 영향을 평가하기 위해 눈 주변, 코, 입 부분으로 나누어 합성 결과를 보였다. Face Parsing 방법을 함께 사용하였을 때 얼굴의 주요 영역인 입과 눈 모양을 보존하는 것을 알 수 있다. 또한 피부 요소 3가지에 대해 실제 영상의 피부 요소와 비슷한 분포로 생성되어 사실적인 영상을 합성한다. 그림 5는 얼굴 전체 영역에 대해 입력 라벨 영상에 따른 합성 결과를 나타낸다. 2904×4354 해상도의 영상을 훈련 때와 동일하게 512×512 해상도로 잘라 테스트하였다. Stride 크기는 32로 설정하고 겹치는 부분은 RGB 평균을 계산하여 자연스럽게 이어지는 전체 영상을 완성하였다. 얼굴 구조가 유지된 채 GT 영상과 비슷하게

피부가 사실적으로 합성됨을 보였다.

정량 평가로써 실제 영상과 합성된 영상의 PSNR과 SSIM 결과를 보여준다. 피부, 눈썹, 코, 윗입술, 아랫입술을 찾는 Face parsing 방법이 얼굴 합성 결과에 미치는 영향을 평가하기 위해 Face parsing 방법 유무에 따른 결과를 비교하였다. 이를 위해 주름, 모공, 홍조, 주름 및 모공, 배경으로 이루어진 라벨 영상을 이용해 훈련하였다. 훈련 환경과 이외 방법은 동일하게 적용하였다. 표 1에서 Face parsing 방법을 활용하였을 때 PSNR과 SSIM 지표에서 더 높은 성능을 보인다.

표 1. 얼굴 합성의 정량적 평가

Table 1. Quantitative evaluation of facial synthesis

Method	PSNR [dB]	SSIM
Ours (w/ Face Parsing)	25.59	0.7295
Ours (w/o Face Parsing)	24.55	0.7221

피부 요소 조절 방법을 사용하였을 때 편집된 라벨 영상들로부터 다양한 얼굴 영상을 합성할 수 있다는 것을 보였



그림 4. 얼굴 영역에 따른 정성적 결과

Fig. 4. Qualitative results of partial facial synthesis

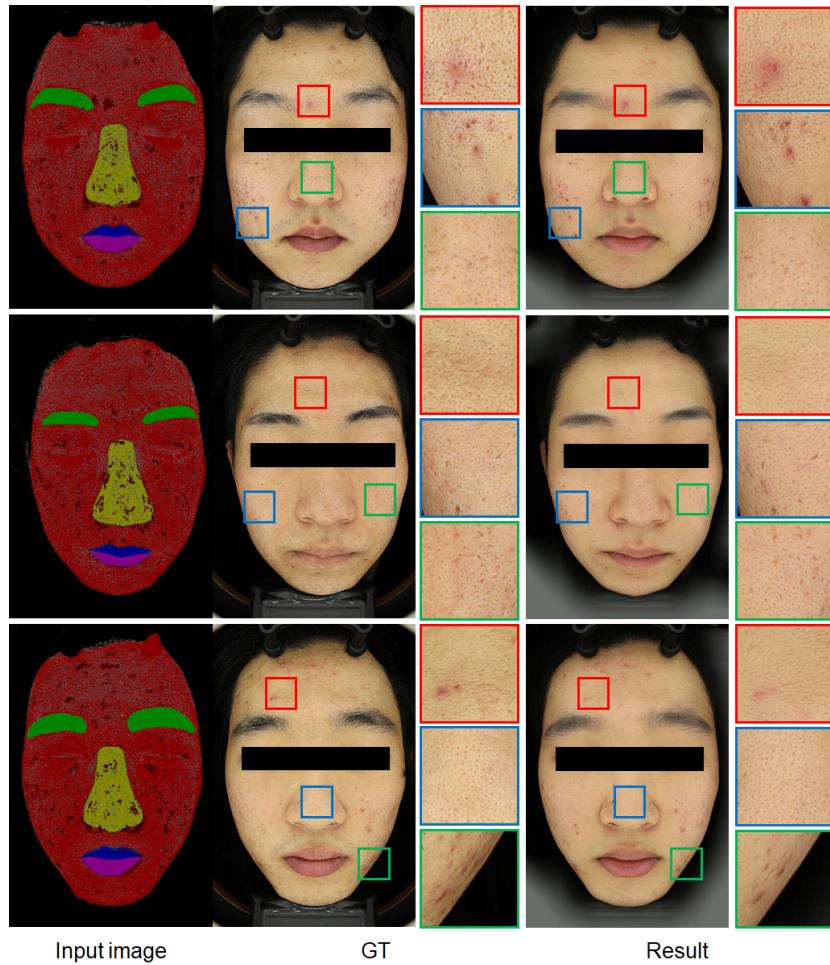


그림 5. 얼굴 합성의 정성적 결과
Fig. 5. Qualitative results of facial synthesis

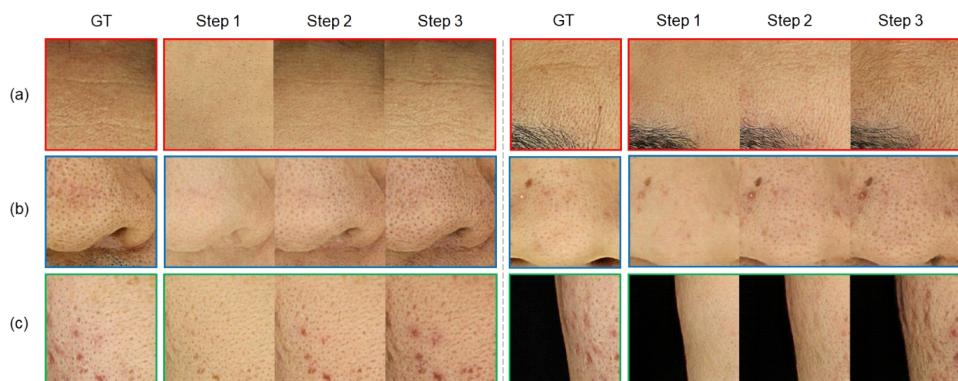


그림 6. 피부 요소 조절에 따른 얼굴 합성 결과. (a) 주름, (b) 모공, (c) 홍조
Fig. 6. Facial synthesis results according to skin element control. (a) Wrinkle, (b) Pore, (c) Redness

다. 라벨 영상을 편집하여 각 피부 요소가 감소 혹은 증가했을 때 영상을 그림 6을 통해 나타냈다. 조절 단계를 3단계로 나누어 오른쪽으로 갈수록 단계가 올라가며 그에 따라 각 피부 요소를 증가시킨 결과이다. 1단계는 기준에 존재하는 피부 요소를 감소시키는 단계이다. 피부 요소가 존재하는 픽셀의 라벨 종류를 변경하는데, 피부 또는 피부 요소의 가중치 비율을 2:1로 설정하여 무작위로 값을 설정한다. 2단계와 3단계는 기준 피부 요소가 존재하는 픽셀 주위의 위, 아래, 왼쪽, 오른쪽 픽셀의 라벨 종류를 변경하였다. 피부 요소와 픽셀의 기준 라벨 종류의 가중치 비율을 2단계는 4:1, 3단계는 6:1로 설정하여 합성한 결과이다. 합성 결과를 통해 얼굴 형태는 그대로 유지되면서 피부 요소가 달라지는 것을 확인할 수 있다. 픽셀의 라벨 종류를 설정하는 가중치 비율과 단계를 설정하여 다양한 유형의 피부를 표현할 수 있다.

V. 결 론

본 논문은 피부 분석기에서 취득한 피부 요소 영상과 얼굴 영상을 이용해 피부 미세요소 조절을 통한 얼굴 합성 기법을 제안한다. 기존 연구에서 다루지 못했던 얼굴의 미세 피부요소 데이터를 취득하고 합성하는 연구로서 의의가 있다. 또한 라벨 영상 생성 방법으로 얼굴 영역의 라벨을 함께 이용함으로써 정성적 및 정량적 평가 결과 비교를 통해 우수함을 보였다. 더 나아가 이미지 프로세싱 방식과 심층신경망 기반의 방식을 융합한 기법으로 라벨 영상을 피부 요소 존재 확률을 조절하여 여러 라벨 영상을 생성함으로써 이에 대응하는 다양한 얼굴 영상 합성을 가능하게 한다. 기존 얼굴 합성 연구에서 피부 수준의 합성으로 발전시킨 연구로 이와 관련된 의료와 미용 분야에 활용할 수 있다. 또한, 여드름과 색소침착과 같은 다른 피부 요소로 확장하여 연구를 진행할 계획이다.

참 고 문 헌 (References)

- [1] Mark-Vu, https://www.psiplus.co.kr/page/?M2_IDX=8160 (accessed Mar. 15, 2022).
- [2] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Proc. Advances in Neural Information Processing Systems*, December 2014.
- [3] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [4] Y. Choi, M. Choi, M. Kim, J. -W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.
doi: <https://doi.org/10.1109/cvpr.2018.00916>
- [5] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2020.
doi: <https://doi.org/10.1109/cvpr42600.2020.00926>
- [6] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: A StyleGAN encoder for image-to-image translation," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2021.
doi: <https://doi.org/10.1109/cvpr46437.2021.00232>
- [7] Z. Huang, S. Chen, J. Zhang, and H. Shan, "PFA-GAN: Progressive face aging with generative adversarial network," *IEEE Trans. on Information Forensics and Security*, Vol. 16, pp. 2031-2045, December 2020.
doi: <https://doi.org/10.1109/tifs.2020.3047753>
- [8] T. -C. Wang, M. -Y. Liu, J. -Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.
doi: <https://doi.org/10.1109/cvpr.2018.00917>
- [9] P. Isola, J. -Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2017.
doi: <https://doi.org/10.1109/cvpr.2017.632>
- [10] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," *arXiv preprint arXiv:1803.07422*, 2018.
- [11] J. -Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc. IEEE International Conference on Computer Vision*, October 2017.
doi: <https://doi.org/10.1109/iccv.2017.244>
- [12] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," *Proc. International Conference on Learning Representations*, April 2018.
- [13] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2020.
doi: <https://doi.org/10.1109/cvpr42600.2020.00813>
- [14] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Proc. Advances in Neural Information Processing Systems*, December 2020.

- [15] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Alvarez, “Invertible conditional GANs for image editing,” arXiv preprint arXiv:1611.06355, 2016.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, Vol. 86, No. 11, pp. 2278-2324, November 1998.
doi: <https://doi.org/10.1109/5.726791>
- [17] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” Proc. IEEE International Conference on Computer Vision, December 2015.
doi: <https://doi.org/10.1109/iccv.2015.425>
- [18] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “AttGAN: Facial attribute editing by only changing what you want,” IEEE Trans. on Image Processing, Vol. 28, No. 11, pp. 5464-5478, May 2019.
doi: <https://doi.org/10.1109/tip.2019.2916751>
- [19] C. - H. Lee, Z. Liu, L. Wu, and P. Luo, “MaskGAN: Towards diverse and interactive facial image manipulation,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2020.
doi: <https://doi.org/10.1109/cvpr42600.2020.00559>
- [20] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “Bisenet: Bilateral segmentation network for real-time semantic segmentation,” Proc. European Conference on Computer Vision, September 2018.
doi: https://doi.org/10.1007/978-3-030-01261-8_20

저자 소개



김 유 진

- 2020년 8월 : 인하대학교 정보통신공학과 학사
- 2020년 9월 ~ 현재 : 인하대학교 전기컴퓨터공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-1990-2704>
- 주관심분야 : 컴퓨터비전, deep learning, face synthesis



박 인 규

- 1995년 2월 : 서울대학교 제어계측공학과 학사
- 1997년 2월 : 서울대학교 제어계측공학과 석사
- 2001년 8월 : 서울대학교 전기컴퓨터공학부 박사
- 2001년 9월 ~ 2004년 2월 : 삼성중합기술원 멀티미디어랩 전문연구원
- 2007년 1월 ~ 2008년 2월 : Mitsubishi Electric Research Laboratories (MERL) 방문연구원
- 2014년 9월 ~ 2015년 8월 : MIT Media Lab 방문부교수
- 2018년 7월 ~ 2019년 6월 : University of California, San Diego (UCSD) 방문학자
- 2004년 3월 ~ 현재 : 인하대학교 정보통신공학과 교수
- 2020년 4월 ~ 현재 : 인하대학교 인공지능융합연구센터 센터장
- ORCID : <https://orcid.org/0000-0003-4774-7841>
- 주관심분야 : 컴퓨터비전 및 그래픽스, deep learning, GPGPU