

피부 미세 요소가 포함된 분할 맵을 이용한 얼굴 영상 합성 모델 분석

김유진, 박인규

인하대학교 전기컴퓨터공학과

kimyujin1088@gmail.com pik@inha.ac.kr

Analysis of Facial Image Synthesis Models using Segmentation Maps including Skin Microelements

Yujin Kim, In Kyu Park

Department of Electrical and Computer Engineering, Inha University

요약

분할 맵 기반 얼굴 합성 분야의 기존 연구들은 피부 영역을 하나의 라벨로 취급한다. 이는 피부 내 미세한 요소를 표현하지 못하며 고해상도 영상 합성 성능 부족이라는 결과를 초래한다. 본 논문에서는 이러한 문제점을 극복하고자 분할 맵에 주름, 모공, 홍조와 같은 피부 요소를 추가하여 이로부터 얼굴 영상을 합성하는 기법을 제안한다. 기존 분할 맵 기반 영상 합성 연구에 제안하는 기법을 적용하여 정량적 및 정성적 비교를 통해 성능이 개선됨을 보인다.

1. 서론

얼굴 합성은 나이를 변화시키거나 다른 사람 얼굴로 교체하는 용도와 같이 다양하게 활용된다. 생산적 적대 신경망을 이용해 스타일 코드를 직접 조작하거나 분할 맵 또는 텍스트 기반으로 영상을 합성하는 방법이 있다. 분할 맵을 기반으로 하는 얼굴 합성 분야는 눈, 코, 입, 귀와 같은 얼굴 내 영역을 분할하여 분할 맵을 만들고 이를 이용해 원하는 영역을 편집하여 영상을 합성하는 연구이다. 피부가 가장 큰 영역을 차지함에도 불구하고 기존 연구들은 이를 하나의 라벨로 분할하여 맵을 생성하여 학습한다. 이와 같은 방법은 사실적인 얼굴 합성을 가능하게 하나 전체적인 얼굴 영역을 합성하는 데 특화되어 주름, 모공과 같은 피부 미세 요소들을 표현하는 데 성능이 부족하고 고해상도 얼굴 합성 영상에 취약하다.

본 논문에서는 얼굴 분할 맵에 피부 요소를 추가하여 기존 연구들의 성능을 향상시킬 수 있는 기법을 제안한다. 기존 연구에서 사용되는 얼굴 분할 맵에 피부 분석기를 이용해 취득한 피부 요소 라벨을 추가한다. 이 과정을 통해 생성된 데이터셋을 이용해 대표적인 분할 맵 기반 영상 합성 모델에 제안하는 기법을 적용하였을 때 정성적 및 정량적으로 성능이 향상됨을 보인다. 또한 피부 미세 요소 표현과 고해상도 얼굴 합성이 가능함을 보여준다.

2. 본론

2.1 데이터 취득

제안하는 기법은 얼굴의 영역뿐만 아니라 피부 내 미세 요소가 추가된 분할 맵을 이용한다. 얼굴 RGB 영상과 그에 대응되는 피부 요소 영

을 취득하기 위해 Mark-Vu 피부 분석기[1]를 이용하였다. 얼굴 RGB 영상으로부터 피부 요소를 추출하기 위해서는 전문적인 지식이 필요하며 많은 시간과 비용이 소요되기 때문에 피부 분석기를 이용해 데이터셋을 취득하였다. 남자 91명과 여자 93명을 대상으로 2904 × 4354 해상도의 얼굴 정면 영상을 취득하였다. 일반광에서 취득한 영상으로부터 주름과 모공 영상을 취득하였으며 편광 영상으로부터 홍조 영상을 취득하였다.

2.2 얼굴 합성 모델 학습

얼굴 RGB 영상에 대한 분할 맵을 생성하기 위해 기본 얼굴 영역을 분할해주는 학습된 모델과 피부 분석기로 취득한 피부 요소 분할 맵을 이용하였다. 먼저 CelebAMask-HQ 데이터셋[2]으로 훈련된 BiseNet[3] 기반 face parsing 모델을 이용해 피부, 눈썹, 코, 윗입술, 아랫입술 영역을 추출한다. 피부 분석기로 취득한 피부 요소 중 주름, 모공, 홍조를 추가하여 총 10가지 라벨을 가진 분할 맵을 생성한다. 라벨 종류는 배경, 피부, 눈썹, 코, 윗입술, 아랫입술, 주름, 모공, 홍조, 주름 및 홍조이다. 한 픽셀에 주름과 홍조가 동시에 존재하는 경우를 처리하기 위해 주름 및 홍조 라벨을 추가하였다.

제안하는 기법을 검증하기 위해 영상 합성 모델을 선정하였다. 대표적인 분할 맵 기반 영상 합성 모델인 Pix2PixHD[4], SPADE[5], SEAN[6]에 대해 학습을 진행하였다. 각 모델마다 face parsing 모델만을 이용해 생성된 분할 맵과 얼굴 RGB 영상으로 이용하는 경우와 피부 요소까지 추가된 분할 맵과 얼굴 RGB 영상으로 이용하는 경우로 나누어 학습한다. 세 모델 모두 분할 맵 기반으로 얼굴 영상을 합성하며 라벨 개수와 종류를 다양하게 설정하여 학습할 수 있다.

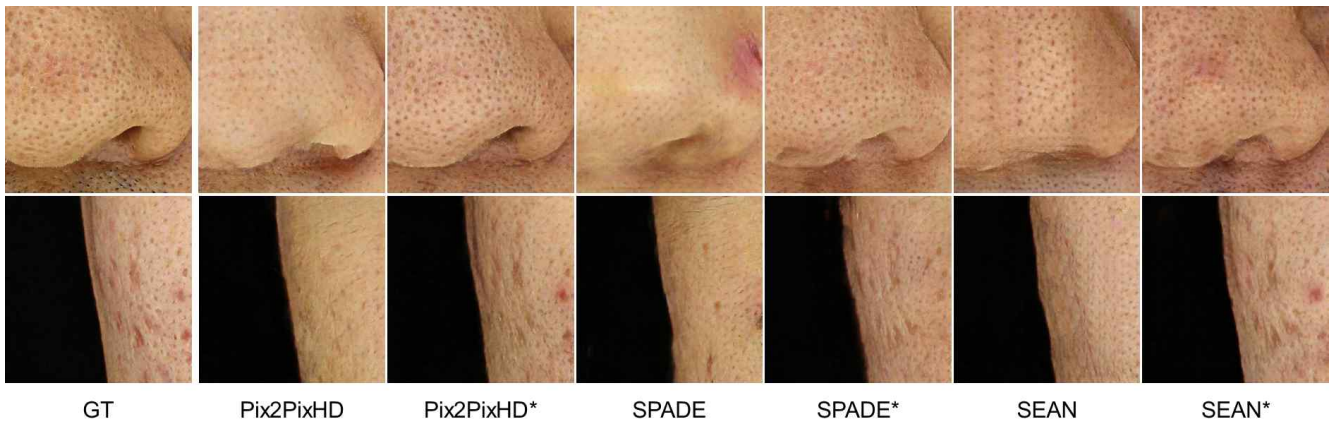


그림 1. 얼굴 합성에 대한 정성적 비교. *은 각 합성 모델에 제안하는 기법 적용 후를 나타냄.

3. 실험 결과

본 논문에서는 Mark-Vu 피부 분석기를 이용해 취득한 데이터셋을 기반으로 학습 및 테스트를 진행하였다. 합성 모델 학습을 위해 4개의 A100 GPU를 이용하였다. 공통적으로 학습률은 0.0002, 배치 사이즈는 16, 옵티마이저는 Adam으로 설정하고 200 epoch동안 학습하였다. 피부 분석기로 취득한 원본 영상에서 512×512 패치로 잘라 총 7,544장을 학습에 사용하였다.

제안하는 기법을 평가하기 위해 피부 분석기로 취득한 얼굴 RGB 영상에 대한 분할 맵을 입력으로 합성 모델을 이용해 얼굴 RGB 영상을 합성하였다. 그림 1을 통해 세 가지 모델에서 제안하는 기법을 적용한 전과 후를 비교하였다. 기존 분할 맵으로 학습하였을 때보다 피부 요소를 추가한 분할 맵으로 학습한 결과가 모공과 홍조와 같은 피부 요소를 잘 표현한다. 또한 코와 같은 얼굴 모양 복원에 우수함을 알 수 있다. 표 1은 512×512 해상도의 920장의 얼굴 패치 영상에 대해 PSNR과 SSIM 값의 평균을 계산한 결과이다. 제안하는 기법 적용 후 정량적으로 개선됨을 보인다. Pix2PixHD 모델의 성능이 가장 많이 개선되었으며 PSNR은 약 3.76, SSIM은 약 0.0963 증가하였다. 적용 후 모델 간 비교 결과는 SEAN 모델이 가장 성능이 좋으며 다음으로 Pix2PixHD, SPADE 순이다.

표 1. 얼굴 합성에 대한 정량적 평가 결과. *은 각 합성 모델에 제안하는 기법 적용 후를 나타냄.

Method	PSNR [dB] ↑	SSIM ↑
Pix2PixHD	21.83	0.6332
Pix2PixHD*	25.59	0.7295
SPADE	21.04	0.5787
SPADE*	23.47	0.7055
SEAN	25.04	0.6635
SEAN*	25.87	0.7457

4. 결론

본 논문에서는 분할 맵 기반 얼굴 합성 모델에 적용할 수 있는 피부 요소가 포함된 얼굴 분할 맵을 이용하는 기법을 제안하였다. 피부의 미

세 요소를 표현하지 못하여 사실적인 피부 표현이 부족하다는 기존 연구의 단점을 극복하고 실험 결과를 통해 정량적 및 정성적으로 우수한 성능을 보인다. 이를 확장하여 피부 요소를 다양하게 조절함으로써 의료와 미용 분야에 활용될 수 있을 것으로 보인다.

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2020-0-01389, 인공지능융합연구센터지원(인하대학교)). 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2022R1A4A1033549).

참고문헌

- [1] M a r k - V u . https://www.psipius.co.kr/page/?M2_IDX=8160.
- [2] C. -H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation," *Proc. IEEE/CVF CVPR*, pp. 5549-5558, June 2020.
- [3] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," *Proc. ECCV*, pp. 325-341, September 2018.
- [4] 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 CVPR*, pp. 8798-8807, June 2018.
- [5] T. Park, M. -Y. Liu, T. -C. Wang, and J. -Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," *Proc. IEEE/CVF CVPR*, pp. 2337-2346, June 2019.
- [6] P. Xhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," *Proc. IEEE/CVF CVPR*, pp. 5104-5113, June 2020.