

## 조각별 아핀변환을 활용한 실시간 표정 재구성

권민지, 김명철, 옥수빈, 이대호  
경희대학교

minmin0916@khu.ac.kr, 2018102095@khu.ac.kr, osb0323@khu.ac.kr, nize@khu.ac.kr

Real-Time Facial Expression Reenactment  
using Piecewise Affine TransformationMinji Kwon, Myoungchul Kim, Subin Ok, Daeho Lee  
Kyung Hee Univ.

## 요 약

본 논문은 학습된 네트워크를 사용하지 않고, 단순한 영상처리 기법인 조각별 아핀변환(piecewise affine transformation)을 사용하여 실시간으로 얼굴 표정을 재구성할 수 있는 새로운 시스템을 제안한다. 이는 기존의 GAN (generative adversarial network) 기반의 선행연구와 달리, 간단한 프로세스로 빠르게 변환된 얼굴 영상을 생성한다.

## I. 서 론

COVID-19 로 인해 수업, 회의와 같이 대면으로 이루어지던 사회활동이 인터넷 상에서 이루어지며 사람들은 강제적으로 메타버스(metaverse)를 경험하고 익숙해졌다. 국내외에서 제페토(zepeto)와 로블록스(roblox)로 잘 알려진 메타버스는 초월이라는 뜻의 메타(meta)와 현실세계를 의미하는 유니버스(universe)의 합성어이다[1]. 현실세계가 가상의 공간으로 확장된 개념으로, 사람들은 가상의 캐릭터에 자신을 투영시켜 감정을 표현할 수 있다. 특히, 감정을 가장 잘 드러낼 수 있는 비언어적 표현인 표정을 캐릭터에 전달함으로써 사용자의 몰입감을 높일 수 있다[2]. 따라서 본 논문에서는 실시간으로 타겟 이미지에 얼굴 표정을 전달하여 재구성할 수 있는 새로운 시스템을 제안한다.

## II. 관련연구

얼굴 표정 재구성(facial expression reenactment)은 소스 영상의 사람의 표정을 타겟 영상의 사람의 얼굴에 전달하여 타겟 영상의 얼굴 표정을 재구성하는 기법을 의미한다. 표 1 과 같이 네트워크 기반의 접근 방식과 컴퓨터 그래픽 기반의 접근 방식이 존재한다.

표 1. 얼굴 표정 재구성관련 선행연구 도표.

Category	Name	Expression Category
GAN	StarGAN[3]	Y
	AttGAN[4]	Y
	UGAN[5]	Y
	FaceGAN[6]	Y
	LandmarkGAN[7]	N
3D shape model	Blanz, V. et al.[8]	Y
	Thies, J. et al.[9]	Y
2D warping	Averbuch-Elor, H. et al.[10]	N

표 1 에서 Expression Category 는 타겟 영상에서 재구성할 수 있는 얼굴 표정의 종류에 제한이 있는 경우 Y 로 표기하였다. GAN (generative adversarial network) 기반의 아키텍처를 사용하여 변환된 얼굴 표정의 영상을 생성한 경우, 표정을 특정 도메인(domain)으로 정의하여 원본 영상의 얼굴을 변환하는 방식을 취하고 있다. LandmarkGAN[7]을 제외한 나머지 GAN 기반 기법은 변환 가능한 표정이 화남, 슬픔 등으로 한정되어 있다. 그러나 LandmarkGAN 은 얼굴의 랜드마크를 활용하여 표정을 전달하고, 이때 생성기(generator)를 사용하여 타겟 얼굴의 이목구비(identity)를 랜드마크에 맞추어 재생성한다[7].

3D 형상 모델을 사용하여 얼굴 표정을 재구성하는 경우, 무표정(neutral expression)의 얼굴과 다른 종류의 표정들을 가중치의 선형 결합으로 표현한다. 3D 레이저 스캔을 통해 형상 모델을 구축하고 표정 변화에 따라 가중치의 선형 결합을 활용하여 모델을 재구성하거나[8], BlendShapes 라는 3D 모폴로지(morphology) 모델을 사용하여 BlendShapes 가중치를 활용해 표정을 재구성한다[9].

최근 얼굴 표정 재구성관련 연구분야에서 GAN 기반의 생성기를 활용한 연구가 대다수이지만, 얼굴 랜드마크를 기반으로 2D 와핑(warping)을 사용하여 표정을 재구성하는 기법이 제안된 적 있다[10]. 해당 기법은 자동화된 프로세스로 얼굴 랜드마크를 추출하여, 2D 와핑과 후보 정을 통해 재구성된 타겟 얼굴 영상을 생성한다[10]. 모델 학습을 기반으로 표정을 재구성하는 많은 선행연구들과 다르게, 간단한 영상처리 기법을 활용한 본 논문과 가장 유사한 연구이다. 본 논문에서는 타겟과 소스 영상의 무표정한 얼굴 영상, 소스 영상의 표정 영상을 모두 요구하는 선행 기법과 달리, 한 장의 타겟 영상과 한 장의 소스 영상만을 사용하여 실시간으로 표정 전달이 가능한 새로운 시스템을 제안한다.

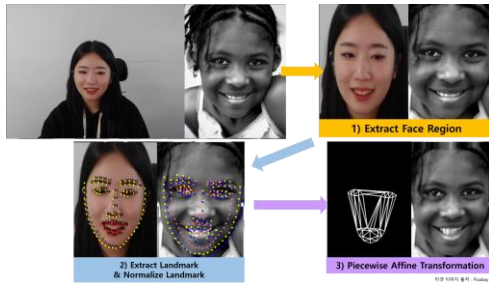


그림 1. 얼굴 표정 재구성 프로세스.

### III. 얼굴 표정 재구성 프로세스

본 논문의 전체적인 얼굴 표정 재구성 프로세스는 그림 1 과 같다. UltraFace[11]을 사용하여 소스 영상과 타겟 영상에서 얼굴 영역을 추출하고, 얼굴 영역에서 PFLD[12]를 사용해 98 개의 랜드마크를 검출한다. 검출된 랜드마크 가운데 가장 큰 X,Y 값과 가장 작은 X,Y 값을 사용하여 소스 영상을 타겟 영상의 랜드마크 영역 크기와 중앙점에 맞추어 정규화 한다. 식 (1)과 같이 랜드마크 영역의 스케일 변화를 조정하기 위한 계수를 계산하고, 식 (2)와 같이 중앙점의 이동을 계산한다. 최종적으로 정규화 된 X 좌표 값은 식 (3)을 통해 얻을 수 있으며, Y 좌표 역시 동일한 과정을 거쳐 정규화 된다.

$$coefficientX = \frac{\max X_{source} - \min X_{source}}{\max X_{target} - \min X_{target}} \quad (1)$$

$$offsetX = centerX_{target} - centerX_{source} * coefficientX \quad (2)$$

$$X_{normalized} = coefficientX * X_{source} + offsetX \quad (3)$$

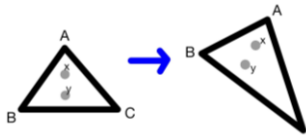


그림 2. 조각별 아핀 변환 예시.

마지막으로 타겟 영상의 랜드마크 영역에 맞추어 정규화 한 소스 영상의 랜드마크를 사용하여 조각별 아핀 변환(piecewise affine transformation)을 수행하였다. 아핀 변환은 점, 직선, 또는 평면을 보존하는 선형 매핑 방법이다. 아핀 변환을 통해 하나의 도형을 다른 위치로 매핑시킬 수 있고, 변환 이후에 모양이 변할 수 있지만 도형 내부의 점과 점 사이의 위치 관계는 보존된다. 그림 2 에서와 같이, 왼쪽 삼각형이 오른쪽 삼각형으로 모양이 변하더라도 내부의 두 점 x, y 사이의 상대적인 위치는 변하지 않는다. 이를 이용하여 사람의 얼굴을 여러 개의 작은 삼각형으로 분할한 다음, 인접한 삼각형들의 위치와 크기를 변경하여 얼굴 표정을 변환할 수 있다. OpenCV의 들로네 삼각분할(Delaunay triangulation)을 사용하여 타겟 영상과 소스 영상의 랜드마크들로 여러 개의 삼각형을 만든다. 타겟 영상의 삼각형을 소스 영상의 삼각형에 맞게 아핀 변환함으로써 표정을 재구성할 수 있다.

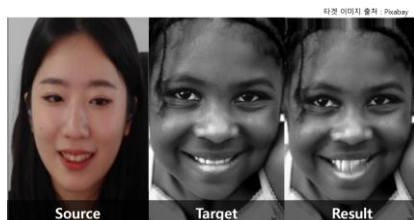


그림 3. 결과 이미지.

### IV. 결과 및 결론

그림 1 과 같이 얼굴 표정 재구성 프로세스를 거친 결과, 그림 3 과 같은 결과 영상을 얻을 수 있다. 본 논문은 학습 기반의 네트워크 사용 없이 간단한 영상처리 기법만을 사용하여 표정 재구성을 수행하였고, CPU 상에서 최고 25fps 를 달성하였다. 빠른 속도로 간단하고 자동화된 프로세스를 거쳐 표정 재구성을 수행할 수 있다는 기여가 있지만, 입을 제외한 얼굴 영역의 표정 변화를 세부적으로 표현하지 못한다는 한계가 있다. 따라서 향후 연구에서는 랜드마크를 통한 표정 검출 단계의 추가적인 보정 단계와, 결과 영상에서 소스 영상의 세밀한 표정 변화를 전달할 수 있는 후보정 단계를 추가할 계획이다.

### ACKNOWLEDGMENT

본 과제(결과물)는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 디지털 신기술 인재양성 혁신공유대학사업의 연구결과입니다.

### 참 고 문 헌

- [1] Lee, B. "The Metaverse World and Our Future," Jour. of the Korea Contents Association, 19(1), pp. 13-17, 2021.
- [2] Kang, M. "Non-verbal Emotional Expressions for Social Presence of Chatbot Interface," Jour. of the Korea Contents Association, 21(1), pp. 1-11, 2021.
- [3] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," IEEE Conf. on computer vision and pattern recognition, pp. 8789-8797, 2018.
- [4] He, Z., Zuo, W., Kan, M., Shan, S., & Chen, X. "Attgan: Facial attribute editing by only changing what you want," IEEE transactions on image processing, 28(11), pp. 5464-5478, 2019.
- [5] Zhu, D., Liu, S., Jiang, W., Gao, C., Wu, T., Wang, Q., & Guo, G. "Ugan: Untraceable gan for multi-domain face translation," arXiv preprint arXiv:1907.11418, 2019.
- [6] Shen, Y., Gu, J., Tang, X., & Zhou, B. "Interpreting the latent space of gans for semantic face editing," IEEE/CVF Conf. on computer vision and pattern recognition, pp. 9243-9252, 2020.
- [7] Sun, P., Li, Y., Qi, H., & Lyu, S. "Landmarkgan: Synthesizing faces from landmarks," Pattern Recognition Letters, 161, pp. 90-98, 2022.
- [8] Blanz, V., Basso, C., Poggio, T., & Vetter, T. "Reanimating faces in images and video," Computer graphics forum, 22(3), pp. 641-650, 2003.
- [9] Thies, J., Zollhöfer, M., Nießner, M., Valgaerts, L., Stamminger, M., & Theobalt, C. "Real-time expression transfer for facial reenactment," ACM Trans. Graph., 34(6), pp. 183-1, 2015.
- [10] Averbuch-Elor, H., Cohen-Or, D., Kopf, J., & Cohen, M. F. "Bringing portraits to life," ACM Trans. on Graph., 36(6), pp. 1-13, 2017.
- [11] Linzaer, "Ultra-Light-Fast-Generic-Face-Detector-1MB," 2021, (<https://github.com/Linzaer/Ultra-Light-Fast-Generic-Face-Detector-1MB>).
- [12] Guo, X., Li, S., Yu, J., Zhang, J., Ma, J., Ma, L., ... & Ling, H. "PFLD: A practical facial landmark detector," arXiv preprint arXiv:1902.10859, 2019.