

Parametric Control of Portrait Images

지도교수 : 김선주 교수님

지도조교 : 전수빈 조교님

Team : GrassC

팀원

2016147514 정윤석

2018147581 전영주

2017126026 이소정

Contents

1. 연구 주제
2. 연구의 필요성
 - a) parametric control 의 필요성
 - b) 얼굴 외 부분 합성
 - c) 기존 연구의 한계점
3. 연구 내용
4. 연구 과정
5. 연구 결과
6. 팀의 구성 및 팀원의 역할
7. 참고문헌

1. 연구주제

- Parametric Control of Portrait Images

입력으로 주어지는 한 장의 사람 얼굴 사진을 3d model로 변환하여 pose와 expression을 parametric control 하는 방법 제안 및 인터페이스 설계

2. 연구의 필요성

a) parametric control의 필요성

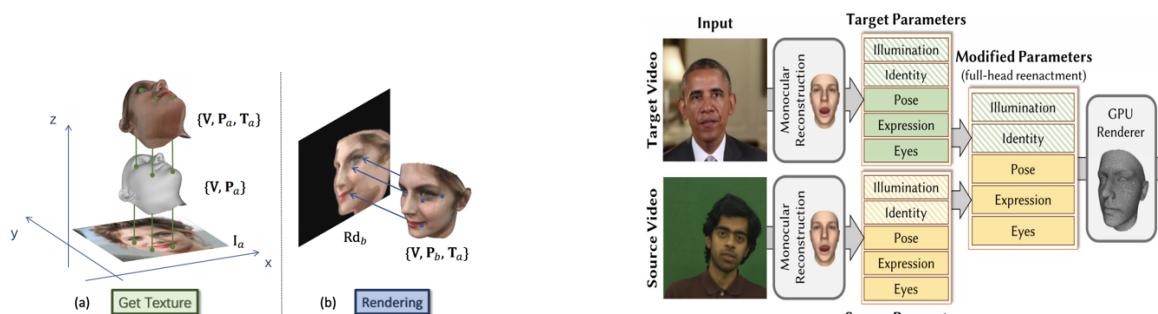
얼굴 편집(Face editing)의 수요가 기하급수적으로 증가하고 있으며, face reenactment 기술은 합성 매체(synthetic media)의 핵심기술로 성장하고 있다. 얼굴 편집의 중요성이 높아짐에 따라 하나의 이미지 파일만을 가지고 parametric control을 통해 얼굴 편집이 가능하다는 것이 이번 연구를 통해 확인된다면 여러 분야에서 좀 더 손쉽게 얼굴 편집이 가능해질 것이다. 그러나, 대부분의 기존 기술은 직관적이고 세밀한 조절을 제공하고 있지 않거나 단일 parameter를 통한 대략적인 편집만이 가능하다. (PIE: Portrait Image Embedding for Semantic Control)

b) 얼굴 외 부분 합성

기존의 일부 방법 중에는 얼굴 부분만을 가지고 3d model을 만드는 기법이 존재한다. 하지만 상업적으로 이용이 가능하게 하려면 좀 더 자연스럽게 표현이 가능해야 하며 그렇기 때문에 얼굴 외 부분까지 합성한 결과물을 만드려고 한다.

c) 기존 연구의 한계점

- Graphics pipeline



(Rotate-and-Render: Unsupervised Photorealistic Face Rotation from Single-View Images
(CVPR 2020)

Deep Video Portraits (Siggraph 2018)

Graphics pipeline은 3차원 이미지를 2차원 래스터 이미지로 표현을 하기 위한 단계적인 방법을 말한다. Graphics pipeline의 경우 구축된 3d model을 사용하기 때문에 감정에 대한 parametric control이 하기에 용이하다. 하지만 머리카락이나 어깨부분 같은 얼굴 외 부분에 대한 이미지 생성이 어렵다.

- GAN 기법을 이용한 face reenactment



(Few-Shot Adversarial Learning of Realistic Neural Talking Head Models (ICCV 2019))

Gan 기법을 사용한다면 3d 모델링 기법 없이 더 넓은 영역에 대해서(머리카락) 사실적인 이미지 합성이 가능하다. 하지만 이미지 생성 과정에서 3d 모델링 기법 없이 진행을 하였기 때문에 기하학적인 추론 없이 이루어진다. 그렇기 때문에 앞의 연구와 같이 3d 모델을 만들 때의 expression parameter에 대해 접근이 불가능하기 때문에 추가적인 parametric control이 불가능하다는 한계가 있다.

3. 연구 내용

- 기존에 있는 모델인 3DMM(BFM)을 사용하여 3D Face Reconstruction을 진행
- pretrained model을 사용하여 각 expression을 대표할 수 있는 image에서 expression parameter를 추출한 후, 3d face에 적용
- Depth map 기법을 사용하여 hair와 ear와 같은 face 이외의 mask 부분 생성
- BFM으로 얻은 face 부분과 mask를 depth net을 이용하여 자연스럽게 합성
- UI 설계

4. 연구 과정

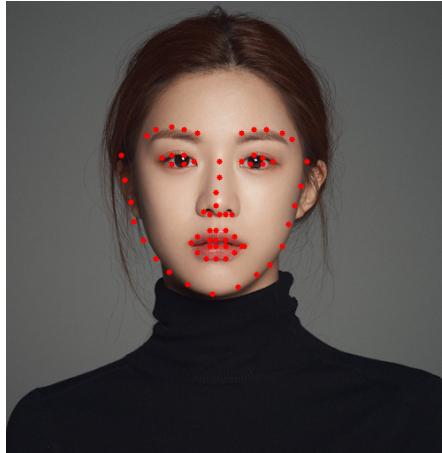
- 임의의 사진을 이용하기 위한 작업

1) Facial landmark 추출

Facial landmark는 얼굴을 표현하는 68개의 점을 말하며,

Landmark를 2D로 추출하여 이미지에 찍으면 다음 사진과 같이 나타낼 수 있습니다.

이 추출한 facial landmark 들은 이미지를 align하고 얼굴 부분만을 crop 해서 BFM 에 사용하기 위한 것입니다.



2) Parsed segmentation 추출

저희가 원하는 모델은 머리카락, 양쪽 귀와 입을 포함한 얼굴 부분을 추출해야 합니다. 기반 논문에서는 이런 부분들을 구현하기 위해 Face Parsing with ROI-Tanh-Warping 논문에서 언급한 ROI-tanh warping 기법을 이용해야하는데, 코드가 공개되어있지 않았습니다. 따라서 공식 리포(<https://github.com/sicxu/Deep3dPortrait>)에서 언급된 모델 (<https://github.com/zllrunning/face-parsing.PyTorch>)을 적용하였으나 전에는 규격 차이가 나서 이 모델을 사용하기 어렵다는 결론에 도달하였습니다. 원래의 모델에서는 피부, 양쪽 눈썹, 양쪽 눈, 코, 입, 윗아래 입술, 목, 옷, 장신구 등 불필요한 부분까지 굉장히 세부적으로 segmentation이 이루어져서 규격이 맞지 않았습니다. 하지만, 저희의 규격에 맞게 머리카락과 그 외 얼굴 부분으로 segmentation하는 것으로 변형시켜서 해결하였습니다.



- Face Reconstruction 진행

저희는 Face reconstruction 을 진행하였는데, 여기에는 기존에 있는 모델인 3DMM 중 Basel Face Model 을 사용하였습니다. 3DMM이란 3D morphable model의 약자로서 주어진 사진으로부터 3차원 얼굴을 복원하는 기술입니다.

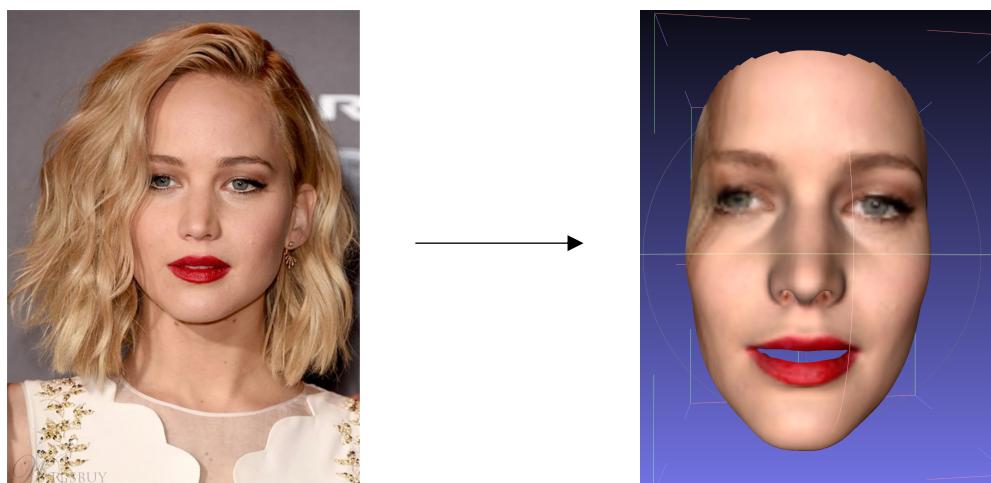
3DMM에서는 얼굴의 모양과 텍스쳐가 다음과 같은 식으로 이루어집니다.

$$\mathbf{F} = \mathbf{F}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bar{\mathbf{F}} + \mathbf{B}_{id}\boldsymbol{\alpha} + \mathbf{B}_{exp}\boldsymbol{\beta}$$

$$\mathbf{T} = \mathbf{T}(\boldsymbol{\delta}) = \bar{\mathbf{T}} + \mathbf{B}_t\boldsymbol{\delta}$$

- F : Face Shape, T : Texture, $\bar{\mathbf{F}}$: Average Face Shape, $\bar{\mathbf{T}}$: Average Texture
- \mathbf{B}_{id} , \mathbf{B}_{exp} , \mathbf{B}_t : 각각 identity, expression, texture 의 PCA bases
- $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$: \mathbf{B}_{id} , \mathbf{B}_{exp} , \mathbf{B}_t 에 대응되는 coefficient vector
- $\bar{\mathbf{F}}, \bar{\mathbf{T}}, \mathbf{B}_{id}, \mathbf{B}_t$: BFM 모델 사용
- \mathbf{B}_{exp} : FaceWarehouse 사용

여기서 F는 얼굴 모양 (Face Shape)을, T는 텍스쳐 (Texture)를 나타냅니다. F라는 average face shape을, T라는 average texture 를 뜻하고, \mathbf{B}_{id} , \mathbf{B}_{exp} , \mathbf{B}_t 는 각각 identity, expression, texture 의 PCA bases를 의미하고, 알파, 베타, 델타는 각각에 대응되는 coefficient vector 에 해당합니다. F바, T바, \mathbf{B}_{id} , \mathbf{B}_t 를 구하기 위해서 저희는 BFM 모델을 사용하였고, \mathbf{B}_{exp} 를 구하기 위해서는 2014년 IEEE에 발표된 FaceWarehouse를 사용하였습니다.



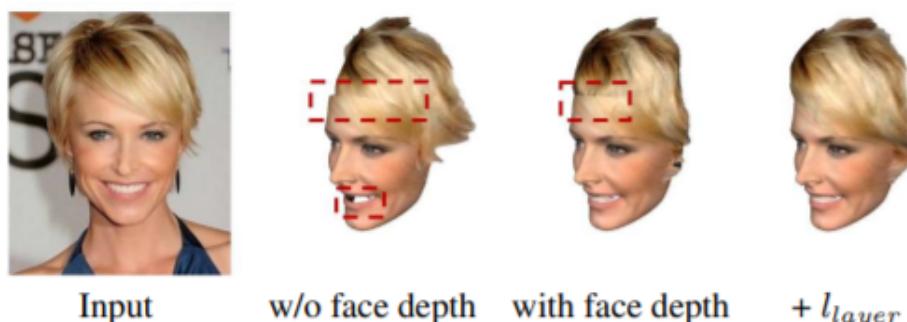
- Segmentation과 Depth map 기법 적용

저희는 다음으로 Segmentation 과 Depth map 기법을 적용하였습니다.

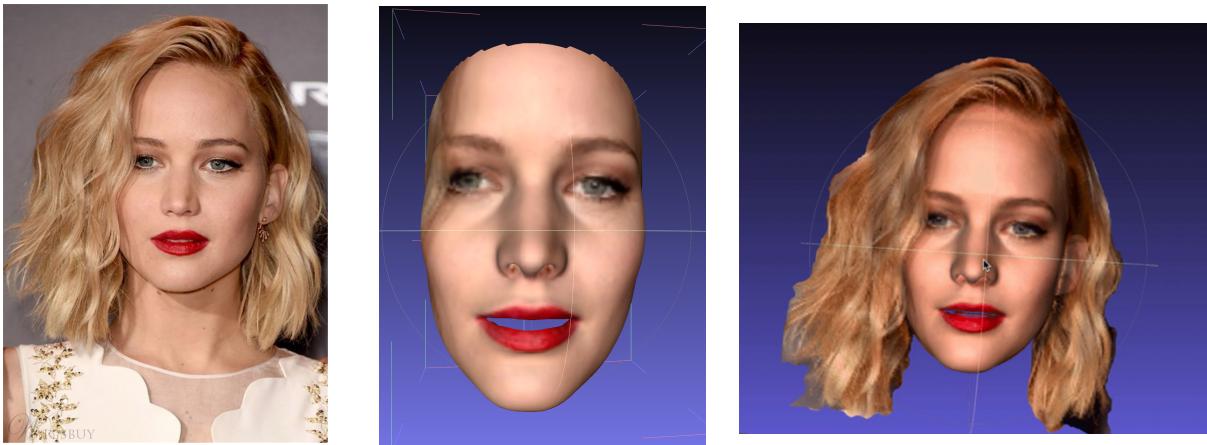
먼저 2d 이미지에 segmentation 을 적용해 hair 와 ear 부분의 mask를 추출합니다.

다음에 BFM으로 얻은 face 부분을 Renderer net을 이용해 depth map을 얻습니다.

이 depth map 과 hair ear 부분의 mask를 depth net 을 이용하여 자연스럽게 합성하여 최종적으로 3d model 을 얻게 됩니다.



위 사진을 보시면 face depth 가 적용이 된 3d model 이 head geometry consistency 와 reconstruction accuracy 에 있어서 더 좋다는 것을 확인할 수 있습니다. 또한, 맨 오른쪽의 사진은 layer-order loss를 고려한 모델인데, 이는 hair 부분과 hair 로 가려진 부분 사이에 상대적인 위치를 정확하게 하기 위한 loss 입니다. 이를 고려한다면 reconstructed 된 shape 가 더 정확히 나올 수 있습니다.



- Parametric Control

1) Head Pose Control

원래 저희의 최종 연구목표는 parametric control한 3d model을 2d image로 reprojection 하는 것이었습니다.

하지만 시간관계상 3d 모델링만 하기로 하였습니다.

3D model을 2D 이미지로 변환하지 않고 UI 상에서 3D model의 head pose만 조정하면 되기 때문에 object 파일값, 즉 vertex의 좌표 자체를 바꿀 필요가 없어서 UI 상의 camera를 rotation 함으로써 head pose control이 가능하게 하였습니다.

추후에 2D 이미지로 다시 변환해야 하게 된다면 rotation에 따라 object 파일의 값도 변경할 예정입니다.

2) Expression Control

기존에는 input image의 expression PCA base를 분석하여 각 표정과 관련된 유의미한 parameter를 추출하려고 하였으나 Parameter와 expression간의 관계를 직접 알아내기가 어려웠습니다.

$$\mathbf{F} = \mathbf{F}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bar{\mathbf{F}} + \mathbf{B}_{id}\boldsymbol{\alpha} + \mathbf{B}_{exp}\boldsymbol{\beta}$$

$$\mathbf{T} = \mathbf{T}(\boldsymbol{\delta}) = \bar{\mathbf{T}} + \mathbf{B}_t\boldsymbol{\delta}$$

따라서 Input image의 expression PCA bases를 분석하여 조정하는 방법 대신, 각 expression을 대표할 수 있는 image의 expression PCA bases와 연관된 파라미터인 β 를 바꾸는 방법을 사용하였습니다. 파라미터 β 에 해당하는 값은 공식 리포에서 제공하는 pretrained model 을 이용하여 구하였습니다. Expression의 PCA basis는 BFM 모델에서 Face Warehouse를 이용해서 구하였습니다.

다.

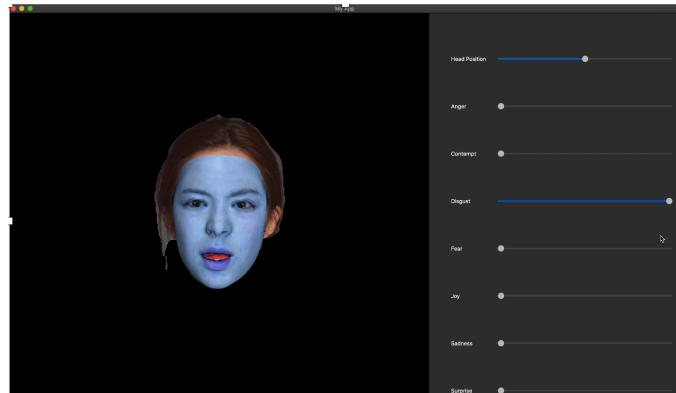
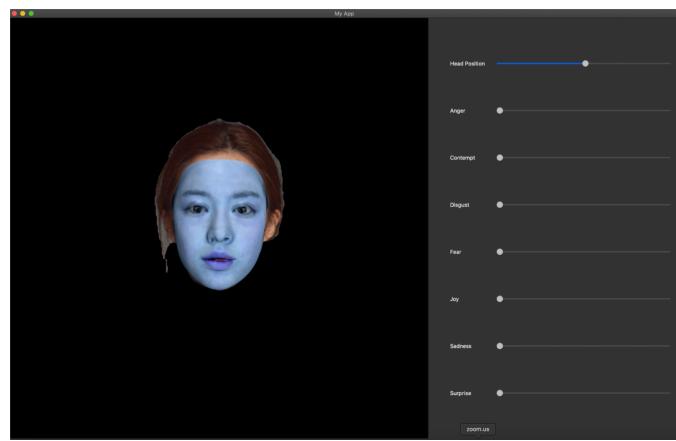
이때 expression은 Joy, Surprise, Contempt, Sadness, Anger, Disgust, Fear로 나누었습니다.

각 expression마다 10단계로 나누었으며, 단계별로 적용하고자 하는 값의 정도를 0~100%까지 10%로 단위로 나누어 input image의 expression을 바꾸는 것에 적용하였습니다.

- UI 설계

UI에 적용하는 input에 관련된 7개의 표정과 0~100%의 각 정도에 해당하는 .obj file을 미리 만들어놓은 다음, pyopengl과 pyqt5를 사용하여 만든 UI와 연동시켰습니다.

5. 연구 결과



6. 팀의 구성 및 팀원의 역할

시기	정윤석	전영주	이소정
2021-3	연구 주제 설정 및 기존 연구 분석		
2021-4	BFM을 사용하여 3D face reconstruction 진행, facial landmark 추 출	BFM을 사용하여 3D face reconstruction 진행, Parsed segmentation 추출	BFM을 사용하여 3D face reconstruction 진행, Depth map 기법을 사 용하여 hair 부분 합성

2021-5	감정에 대한 expression basis의 parameter를 추출하여 target face에 parametric하게 적용, UI 설계 및 모델과 연동
2021-6	프로젝트 전시 및 마무리

7. 참고문헌

- 기반 논문

Sicheng Xu, Jiaolong Yang, Dong Chen, Fang Wen, Yu Deng, Yunde Jia, Xin Tong. 2020. Deep 3D Portrait from a Single Image. In Conference on Computer Vision and Pattern Recognition (CVPR).

- 그 외 논문

Ayush Tewari, Mohamed Elgharib, Mallikarjun B R., Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhöfer, Christian Theobalt. 20 Sep 2020. PIE: Portrait Image Embedding for Semantic Control. SIGGRAPH Asia 2020.

Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, Victor Lempitsky. 2019. Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. In The IEEE International Conference on Computer Vision (ICCV).

Hang Zhou, Jihao Liu, Ziwei Liu, Yu Liu, Xiaogang Wang. 2020. Rotate-and-Render: Unsupervised Photorealistic Face Rotation from Single-View Images. In Conference on Computer Vision and Pattern Recognition (CVPR).

Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Nießner, Patrick Pérez, Christian Richardt, Michael Zollhöfe, Christian Theobalt. 2018. Deep Video Portraits. In the SIGGRAPH.

Jinpeng Lin, Hao Yang, Dong Chen, Ming Zeng, Fang Wen, Lu Yuan. 2019. Face Parsing with RoI Tanh-Warping. In Conference on Computer Vision and Pattern Recognition (CVPR).