

가상 온라인 미팅 시스템을 위한 실감 3D 아바타 생성

*장성걸 **김주은 *김형민 *박종일¹

*(주)아리아엣지, **한양대학교

*sg.zhang@aria-edge.com **wonie08@hanyang.ac.kr

*hm.kim@aria-edge.com *jongil.park@aria-edge.com

Realistic 3d avatar generation for virtual online meeting system

*Zhang, Xingjie **Kim, Jueun *Kim, Hyungmin *Park, Jong-Il

*ARIA-Edge, **Hanyang University

요약

지속되는 팬데믹과 함께 비대면 시대가 개막 되었다. 따라서 사람과 사람사이 접촉을 최소화 하면서, 동시에 생활과 작업의 효율을 보장하는 시스템에 대한 수요가 사회의 이슈로 떠오르게 되었다. 이런 시대적 배경에 발 맞춰 회사생활 또한 급속한 변화를 이루어내고 있다. 그중에서 컴퓨터 비전과 그래픽스 기술의 눈부신 발전과 함께 온라인과 가상공간에서 업무를 보려고 하는 시도가 좋은 대안으로 세간의 주목을 받고 있다. 본 논문에서는 가상공간에서 미팅을 진행 할 수 있는 가상 온라인 미팅 시스템과 이런 시스템을 구성하는데 제일 중요한 과제인 실감 3차원 아바타를 생성하는 방법을 제안한다.

1. 서론

역사적으로 질병은 항상 인류의 생존을 위협해 왔다. 하지만 현대의 발전된 과학기술은 그 위협을 기하급수적으로 낮추었다. 따라서 많은 전염병들은 하나둘씩 역사 속으로 사라졌다. 최근 들어 사스와 같은 신종 전염병들이 출현하기도 했지만, 이런 질병들은 여전히 전문적인 의료와 방역 분야의 관심사였을 뿐, 이로 인해 우리의 생활방식이 바뀌지는 않았다. 하지만 이번 코로나사태는 순식간에 의료적인 차원의 문제를 넘어 우리의 생활을 강타하고 우리의 삶을 송두리째 바꿔 놓았다. 따라서 기존의 접촉을 기반으로 하는 생활방식에는 큰 제약이 발생하게 되었으며, 결국 비접촉 사회로의 변화를 이끌어 내는 계기가 되었다.

질병에 저항하기 위한 행정적 혹은 자발적 조치들이 연달아 이어졌고 사람과 사이는 코로나라는 무형의 벽에 가로막혀 버렸다. 하지만 인류는 만남과 접촉을 멈추었을 뿐 결코 소통을 멈추지는 않았다. 기존의 기술들을 동원하여 여전히 소통을 이어갔으며, 또 다른 방식으로 사회는 작동하기 시작하였다. 그 중에서 컴퓨터 비전과 그래픽스기술을 기반으로 하는 가상공간에서의 만남이 세간의 주목을 받고 있다. 이런 방식은 아바타를 이용함으로써 단순 음성 통화나 문자를 이용한 채팅에 비해 정보전달이 훨씬 직관적이고 효율적이다. 동시에 영상통화에 비해 실제 자신의 얼굴을 들어내지 않아 개인정보보호 측면에서도 이득이다. 이는 특히 업무미팅이나 강의, 모임 같은 현시대 주요한 생활방식에 적합하여 큰 관심을 받게 되었다.

본 논문에서는 이런 가상공간의 한 예로써 온라인 미팅 시스템을 제안하고, 또 이런 시스템을 구성하는 제일 중요한 요소인 3차원 아바타를 생성하는 방법을 제안한다.

2. 가상 온라인 미팅 시스템

본 논문에서 제안하는 가상 온라인 미팅 시스템은 주요하게 아바타의 생성과 모션 추적 및 재현 두 가지 부분으로 나누어진다. 사용자들은 우선 한 장의 이미지를 이용하여 자신의 아바타를 생성한다. 이렇게 생성된 아바타는 사용자가 미팅에 참여할 경우 사용자를 대표하여 가상공간에 스폰 된다. 다음 카메라를 이용하여 사용자의 얼굴 모션을 분석하고 추적한다. 이때 얻어진 face landmark의 정보와 head pose 정보를 이용하여 가상공간에 있는 아바타에 실시간으로 재현을 한다. (그림 1)

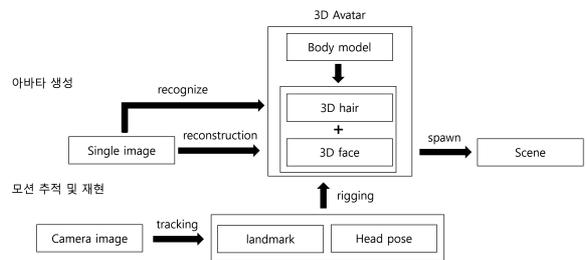


그림 1. System overview

2.1. 실감 3D 아바타 생성 방법

한 장의 이미지를 이용하여 3차원 얼굴 모델을 생성하였다. 우선 BlendFace [6] 기반 모델을 입력된 얼굴 영상과 정합을 시킨다. 본 논문에서는 FLAME [7] 모델을 이용하여 3차원 얼굴을 생성하였다. FLAME 모델은 얼굴의 형태, 표정 및 포즈 3가지 정보를 이용하여 새로운 얼굴을 정의하는데, 이 단계에서는 형태정보만 사용하여 FLAME 모델의 얼굴의 형태를 사용자의 얼굴 형태로 변환 시켰다.

1) 교신저자

다음 생성된 3차원 얼굴 모델에 헤어스타일을 장착시켜줘야 한다. 본 논문에서는 우선 아바타 생성에 사용된 영상 속 인물의 헤어스타일의 종류를 인식한 후, 사전에 준비된 헤어스타일중 그 결과에 대응되는 헤어를 찾아내어 생성된 3차원 얼굴과 정합 시켰다.

마지막으로 이렇게 생성된 3차원 얼굴과 준비된 3차원 몸체 모델을 융합하여 회의에 참석할 아바타를 생성한다.

2.2. 표정 추적 및 재현

3차원 아바타가 준비되면 카메라로부터 입력받은 사용자의 표정과 포즈 정보를 3차원 아바타가 재현 할 수 있게 카메라를 이용하여 실시간으로 표정과 포즈를 추적한다. 3차원 얼굴의 생성에 사용된 FLAME 모델은 한 장의 이미지로부터 표정 파라미터를 계산 할 수 있지만 계산량이 많아 동영상에 적용하기에는 부담이 매우 크다. 따라서 본 논문에서는 실시간 표정추적에 앞서, 우선 sparse coding [5] 시스템을 이용하여 카메라로부터 입력된 face landmark의 분포와 모델의 표정 파라미터 사이 관계를 학습한다. 여기서 학습에 필요한 landmark를 추출하기 위해 FAN [3] 을 사용하였다.

이렇게 학습된 sparse coding 시스템을 이용하여 표정 재현을 진행한다. 이때 sparse coding 시스템과 head pose를 추정하기 위한 알고리즘은 모두 face landmark를 입력으로 사용하게 된다. 동시에 실시간성을 보장하기 위해 face landmark 추출 알고리즘을 보다 빠른 Random forest 기반의 ERT [2] 알고리즘을 채택하여 사용하였다. 그리고 전처리에 사용된 얼굴영역 검출 알고리즘은 S3FD [1] 를 사용하였고 재현 단계에서는 MOSSE(Minimum Output Sum of Squared Error) Object tracking [4] 알고리즘을 이용하여 검출된 얼굴영역을 추적 하였다. 그림 2 에서 표정 추적의 과정을 보여주고 있다.

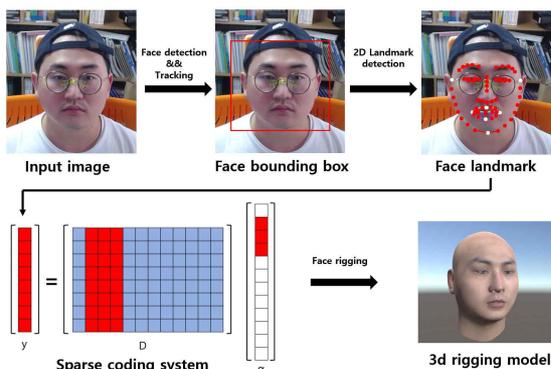


그림 2. 표정 추적 및 재현

3. 실험결과

제안하는 시스템을 시연하기 위해 두 명의 데모 참여자가 각각 자신의 사진을 이용하여 3차원 아바타를 생성하였다. 다음 웹캠을 통해 실시간으로 입력된 영상에서 face landmark와 head pose를 계산하였다. 그 중에서 head pose는 직접 아바타의 head pose로 적용되고, face landmark는 사전에 학습된 sparse coding system을 통해 아바타의 표정 파라미터를 계산 하고 새로운 표정을 생성한다.

그림 3에서 본 논문에서 제안하는 가상 온라인 미팅 시스템의 실제 실행 장면을 보여 주고 있다.



그림 3. 가상 온라인 미팅 시스템

4. 결론

본 논문에서는 비대면 미팅을 위한 가상 온라인 미팅 시스템과 3차원 아바타를 생성하는 방법을 제안하였다. 미팅 참여자는 단 한 장의 이미지만 있으면 이 시스템을 이용해 자신의 아바타를 생성하고 가상공간에 입장을 할 수 있다. 동시에 실험을 통해 이렇게 생성된 아바타가 실시간으로 사용자의 표정과 포즈를 재현하는 것을 확인하였다. 하지만 실감 아바타를 구현하기 위해서는 텍스처, 조명, 미세표정 등 부분을 계속하여 보강할 필요가 있다. 우리는 이 작업들을 추후의 연구를 통해 보완하려고 한다.

감사의 글

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 ICT R&D 지원을 받아 수행된 연구임 (No.2021-000917, 식물 성장 영상 정보를 이용한 식물공장 피노믹스 시스템 개발)

참고문헌

- [1] Zhang, Shifeng, et al. "S3fd: Single shot scale-invariant face detector." Proceedings of the IEEE international conference on computer vision. 2017.
- [2] Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [3] Bulat, Adrian, and Georgios Tzimiropoulos. "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)." Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [4] Bolme, David S., et al. "Visual object tracking using adaptive correlation filters." 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010.
- [5] Lee, Honglak, et al. "Efficient sparse coding algorithms." Advances in neural information processing systems 19 (2006).
- [6] Lewis, John P., et al. "Practice and theory of blendshape facial models." Eurographics (State of the Art Reports) 1.8 (2014): 2.
- [7] Li, Tianye, et al. "Learning a model of facial shape and expression from 4D scans." ACM Trans. Graph. 36.6 (2017): 194-1.