특집 메타버스

실감형 원격 미팅을 위한 3차원 실사 아바타 생성 기술

□ *,**장성걸, *,**김형민, ***서병국, *,**박종일 / * ㈜아리아엣지, **한양대학교, ***ETRI

요 약

최근 메타버스가 장안의 화제로 떠오르면서 많은 사람들의 관심이 쏠리고 있다. 회사의 공간을 극단적으로 줄이고 메타버스로 출근하는 IT기업들이 생겨나는가 하면, 커리큘럼에 메타버스를 적용한 실습교육을 도입하는 의대가 생기기도 했다. 이처럼 사회 생활 전반에 걸쳐 새로운 혁신을 주는 미래의 먹거리로 도약하고 있지만, 그 구현에 있어서 여전히 해결해야 할 기술적 과제가 존재한다. 본 고에서는 회의 참가자가 자신과 똑같은 3차원 실사 아바타를 통해 원격 미팅에 참여하고 다른 참가자와 상호작용할 수 있도록 구성되는 실감형 원격 미팅 시스템을 소개한다. 특히 가장 핵심적 요소 기술인 실사 아바타를 구현하는 방법에 대해 구체적으로 설명한다.

I. 머리말

메타버스(Metaverse)는 가상·초월(meta)과 세계·우주(universe)의 합성어로, 3차원 가상 세계를 뜻하며 정치·경제·사회·문화의 전반적인 측면에서 현실과 가상이 모두 공존할 수 있는 생활형·게임형 가상 세계라는 의미로 해석되고 있다. 메타버스라는 용어의 엄밀한 정의는 아직까지 확립되지 않았지만, 일반적으로는 '현실 세계와 같은 사회적·경제적 활동이 통용되는 3차원 가상공간' 정도의 의미로 널리 사용되고 있다[1]. 저명한 공상과학 소설가 Neal Stephenson은 그의 저서 Snow Crash(1992)에서 메타버스는 컴퓨터에 의해서 생성된 세계로서 고글을 통해 그려지고 이어폰을 통해 들을 수 있는 환상적인 공간이라고 하였다.

컴퓨터가 그래픽으로 만들어낸 가상 세계는 우리에게 가장 친숙한 형태의 메타버스로서, 오픈월드형 온라인 롤플레잉 게임에서부터 린든랩에서 개발된 세컨드 라이프와 같은 생활형 가상 세계에 이르기까지 3차원 컴퓨터 그래픽 환경에서 구현되는 커뮤니티를 총칭하는 개념이다. 우리나라에서도 네이버에서 개발한 제페토는 이용자가 2억 명을 돌파하며 대표적인 메타버스로서의 위용을 과시하고 있다.

특히 코로나19 시대의 도래와 함께 비대면 활동의 수요가 급증하면서 메타버스에 대한 연구와 제품 개발이 전례 없는 각광을 받고 있다. 이전의 메타버스 발전은 게임을 통한 체험을 중심으로 현실과의 연결고리를 찾는 방식으로 진행되었다면, 최근에는 당장 현실적인 문제를 해결하기 위한 시급한 사안으로 대두되고 있으며 원격 미팅 솔루션이 그 대표적인 예라고 할 수 있다.

정부, 단체, 회사, 학교 등 어떤 기관이나 조직을 막론하고 일반적으로는 대면 미팅을 통해 정보를 교환하고 공유하는 것이 가장 기본적인 방법이다. 이렇게 지극히 당연하던 일이 코로나로 인해 기본적인 사회활동에 제약이 생기면서 비대면 원격 화상회의에 대한 수요가 수면 위로 부상하게 된 것이다. 그러나 기존의 화상회의 시스템들은 예전에 비해 발전했음에도 불구하고 몰입감이 부족하고 프라이버시 문제 등이 발생하고 있어 이러한 문제를 해결할 사회적 필요성이 대두되고 있다.

본 고에서는 기존 영상/음성만을 공유하는 원격 미팅 솔루션의 한계를 넘어서 실사 아바타를 이용한 상호작용을 통해 실제 공간에서 회의를 하는 듯한 현장감과 몰입감을 제공하고 프라이버시 문제에도 자유로울 수 있는 시스템을 설명한다. 〈그림 1〉에 나타낸 것은 과학기술정보통신부의 비대면 비즈니스 디지털 기술혁신 사업의 일환으로 ㈜아리아엣지가 ㈜픽스트리와 수행하고 있는 실감형 원격 미팅 시스템이다[21]. 이 메타버스형 원격 미팅 시스템에서는 모든 사용자가 실사 아바타를 통해 회의에 참여하며 회의 공간은 완전한 실감형 가상 세계이다.

이러한 시스템을 실현하는 핵심기술은 실감 아바타 생성 및 상호작용 기술이다. 구체적으로는 3차원 얼굴 모델 기반 실사복원 아바타 모델링 기술과 사용자 표정 변화에 실시간으로 대응 가능한 3차원 얼굴모델 변형기술을 들 수 있으며, 본 고에서는 이에 대해 구체적으로 설명하고자 한다.

II. 실사 아바타 생성 및 변형 기술

실감형 원격 화상회의 시스템을 구현할 때 제일 중요한 부분은 실사 아바타 기술이다. 실사 아바타 기술은 얼굴 모델의 3D 복원과 실시간 표정 재현으로 나누어 설명할 수 있다. 얼굴 모델은 사용자의 특성을 나타내는 제일 중요한 수단이며, 그 모델을 자연스럽게 변형하여 만들어내는 표정도 중요한 정보 전달의 요소로 작용한

<그림 1> 기존 원격미팅(왼쪽)과 메타버스형 원격미팅 시스템 개념도(오른쪽)[21]

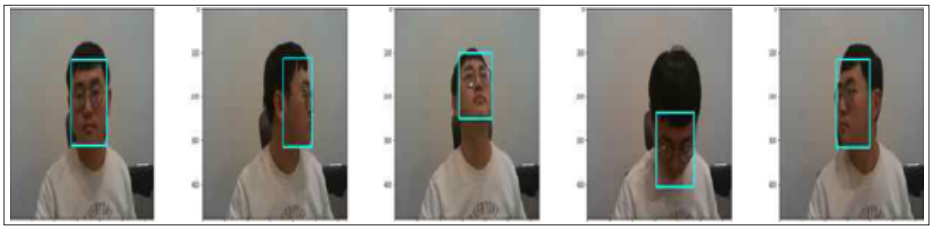
다. 본 장에서는 우선 실감형 아바타를 생성할 때 기본적으로 쓰이게 되는 얼굴 정보 분석 방법을 간략하게 설명하고, 그 다음 얼굴 모델 생성과 실시간 표정 재현 방법에 대해 서술한다.

1. 얼굴 인식 및 안면 포즈 추정 기술

사용자의 2차원 영상으로부터 실감 아바타를 생성하고 표정의 움직임을 추적하기 위해서는 영상 내 얼굴의 몇 가지 기본적인 정보들을 분석하는 것이 필요하다. 그 중에서 얼굴 영역, 얼굴 특징점(Landmark), 안면 포즈 등 정보를 추출하는 알고리즘의 성능이 전반적인 시스템의 성능에 중요한 역할을 하고 있다.

- 얼굴 검출 기술 : 입력된 영상 내에서 얼굴의 위치를 찾는 방법으로서 주로 얼굴 인식이나 얼굴 특징점 검출 및 헤드 포즈 추정 등 알고리즘의 전처리 단계에서 사용된다. 일반적으로 얼굴 위치 추적 알고리즘과 함께 사용되며, 정확하고 안정적으로 얼굴의 위치를 제공하는 것이 핵심이다. 화상회의 시스템에서는 실시간으로 상호작용이 되어야 하기에 최고의 성능을 내는 알고리즘보다 적은 계산량으로도 우수한 성능을 내는 알고리즘을 선택할 필요가 있다. 최근 Single-shot 기반의 딥러닝 모델들이 상대적으로 적은 계산량으로도 좋은 효과를 보여주고 있어, 본 과제에서는 이런 모델을 이용하여 웹캠을 통해 획득한 사용자의 얼굴의 위치를 추정하였다[11].

- 얼굴 특징점 검출 기술 : 얼굴의 기하학적 정보를 나타내는 얼굴 특징점(Landmark)은 얼굴의 포즈를 검출하고 3차원 얼굴 모델을 2차원 이미지(얼굴 텍스쳐)와 정합시키는데 필수적인 요소이다. 얼굴 특징점을 추출하는 알고리즘의 성능에 따라 아바타 생성 시 얼굴 텍스쳐가 뒤틀려 정합되거나, 사용자는 앞을 바라보고 있으나 아바타는 옆을 바라보는



<그림 2> Single-Shot 기반 얼굴 위치 검출 결과

<그림 3> PFLD를 이용한 얼굴 특징점 검출 결과

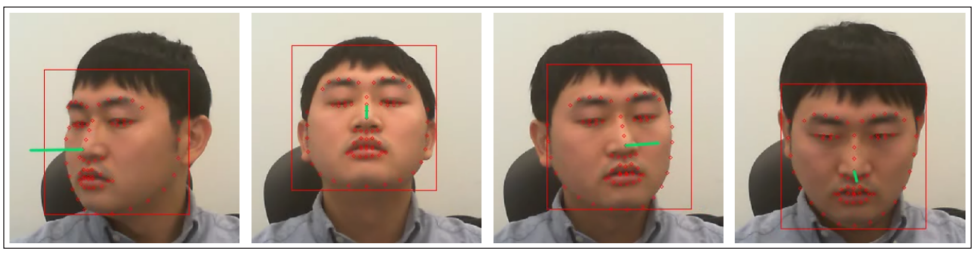
등의 문제가 발생할 수 있다. 또한 주변의 조명 환경이 급격하게 변화하거나 얼굴을 빠르게 움직이는 경우 얼굴 특징점 검출 성능에 영향을 줄 수 있다. 일반적으로 화상회의에서는 야외 환경에서 참여하거나 조명이 급격하게 바뀌는 환경이 아닌 실내 환경에서 사용하는 상황을 상정하여 실내 환경에서 얼굴 포즈 변화에 강인하고 실시간으로 동작 가능한 얼굴 특징점 검출 알고리즘인 PFLD(Practical Facial Landmark Detector)[9]를 이용하였다.

- 얼굴 포즈 추정 기술 : 사용자가 바라보는 방향을 3차원으로 복원된 실사 아바타도 동일하게 같은 방향을 바라볼 수 있도록 하기 위해서는 사용자의 얼굴 포즈를 추정해야 한다. 얼굴 영상에서 사용자가 바라보는 방향을 추정하는 얼굴 포즈 추정은 일반적으로 얼굴 특징점(landmark) 정보에 기반하여 얼굴의 대략적인 3차원 형태를 가정하고 얼굴 포즈를 계산하게 되는데 최근에는 딥러닝 기반으로 직접 얼굴 포즈를 추정하는 방법들이 발표되고 있다[19]. 카메라를 통해 실시간으로 획득한 사용자의 얼굴 포즈 변화를 3차원 실사 복원 아바타에 실시간으로 적용해야 하기에 정확도에서 어느 정도 타협을 보더라도 실시간이 보장되는 POSIT[20] 알고리즘을 이용하여 얼굴 포즈를 계산하였으며, 얼굴 추적 결과를 활용하여 떨림을 보정하였다.

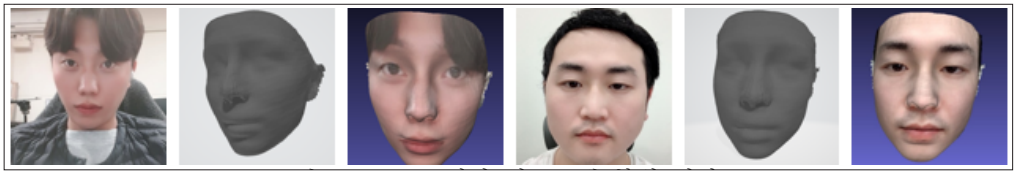
2. 3차원 실사 얼굴모델 복원 기술

메타버스 환경에서 아바타를 이용하여 사용자의 상태를 얼마나 현실감있게 복원하는지는 핵심적인 문제이다. 그 중에서 얼굴은 사람의 캐릭터를 나타내는 제일 중요한 부분으로서 현실감 있는 3차원 얼굴은 참여자들에게 생동감을 선사할 수 있다. 사람들은 상대방의 얼굴을 보면서 이 사람이 누구인지 판단하게 되고, 또 그 사람에 대한 이해는 첫인상으로부터 시작된다고 할 수 있으므로 정확한 실사 얼굴 모델 복원은 매우 중요하다.

- 실사 복원 3D 아바타의 3D 얼굴 모델 복원 : 사용자 얼굴을 3D 모델로 복원하는 일반적인 방법은 3차원 얼굴 기저모델을 이용한 방법이 있다. 사람의 얼굴은 쌍둥이라 할지라도 서로 다른 얼굴을 가지고 있다. 하지만 눈, 코, 입으로 이루어진 전반적인 이목구비는 변하지 않는다. 3차원 얼굴 기저모델은 이런 얼굴들의 평균적인 형태와 얼굴 모양의 차이를 나타내는 기저벡터를 이용하여 어떤 특정한 대상의 얼굴의 형태를 조합해 낸다. 대상 얼굴이 입력되면 우선 얼굴 영상으로부터 랜드마크를 추출한후



<그림 4> 얼굴 포즈 검출 결과



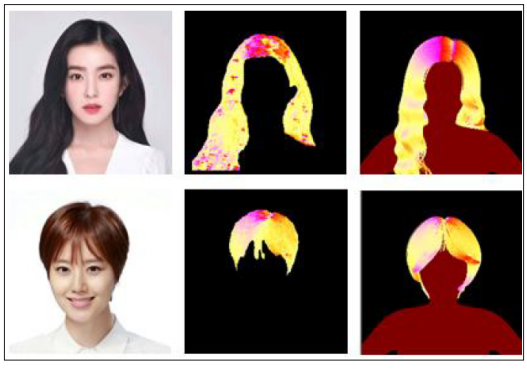
<그림 5> PRNet 기반 얼굴모델 복원 결과

기저벡터들의 가중치를 조절하면서 생성된 3차원 얼굴에서 추출한 랜드마크와 최적의 정합관계를 찾아낸다. 이때 생성된 3차원 얼굴 모델이 입력된 사용자 영상의 얼굴이라고 할 수 있다.

또 다른 방식으로 딥러닝 모델에 기반하는 방식으로 영상으로부터 직접 안면 3차원 포인트 클라우드를 추정하는 방법이 있다. 위 그림은 딥러닝 기반으로 얼굴을 복원한 결과로, PRNet을 이용하여 사용자의 영상으로 얼굴의 3차원 위치 UV map을 생성하고 3차원 얼굴을 복원하였다.

• 실사 복원 3D 아바타의 헤어 모델 복원 : 실사 3D 아바타의 복원에서 또 하나의 중요한 부분은 헤어스타일의 복원이다. 일반적으로 알고리즘을 통해 복원된 3차원 얼굴 모델들은 헤어스타일을 가지고 있지 않다. 따라서 적절한 헤어를 함께 복원해 줄 필요가 있다. 많이 쓰이는 접근방법으로 입력된 영상으로부터 헤어를 직접 복원하는 생성 기반의 헤어 모델 생성 방법과 미리 만들어진 헤어스타일 데이터베이스에서 영상 내 인물의 헤어스타일과 일치하거나 유사한 헤어 모델을 추천하는 인식 기반의 헤어 모델 매칭 방법이 있다.

생성 기반의 헤어 모델 생성 방식은 일반적으로 상당한 계산량이 필요한 복잡한 딥러닝 모델을 사용하여 영상에서 바로 헤어 모델을 생성하는 방식으로 관련 연구들이 활발하게 진행되고 있다. 하지만 실제 제품에 적용하여 시스템화를 하기에는 아직 다소 무리가 있다[15, 16]. 인식 기반의 헤어 모델 매칭 방법은 충분한 헤어 모델 데이터만 가지고 있다면 보다 적은 계산량으로 대응되는 헤어스타일을 찾아낼 수 있을 뿐만 아니라 잘 모델링이 된 헤어 모델을 사용하기에 보다 고품질의 결과를 얻을 수 있다. 최근 모델링 툴들의 발전으로 고품질의 3차원 헤어 모델을 대량으로 획득하는 것이 가능하기에 실사 복원 3D 아바타의 헤어 모델을 제작하여 헤어스타일 복원의 후보군으로 활용하는 것이 직접 영상에서 헤어 모델을 복원하는 것보다 효율적이다. 본 과제에서는 Auto-encoder[17, 18]를 이용하여 헤어스타일 후보군의 특징을 구분할 수 있는 latent를 생성하고 이를 이용하여 헤어 모델의 후보군에 대해 분류를 진행하였다. 또한 동일한 분류 내 비슷한 모델들 사이의 미세한 차이는 인덱스 기반의 특징을 이용하여 인식을 진행하였다[5]. 후보군을 통해 유사한 헤



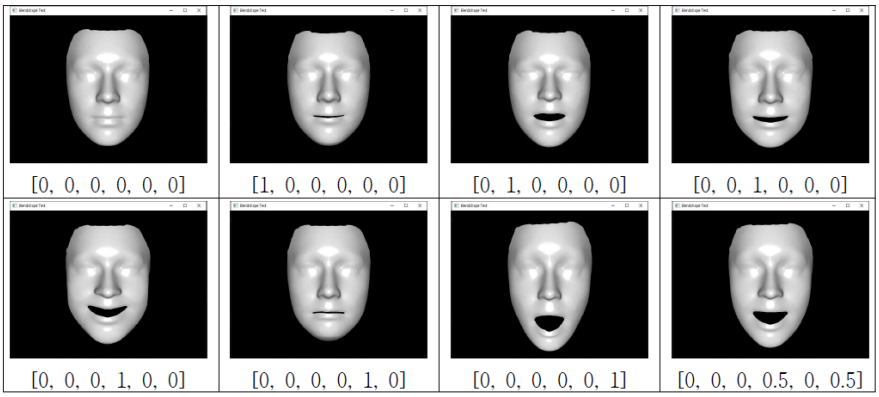
<그림 6> 헤어스타일 인식 결과(좌측 : 사용자 영상, 가운데 : 사용자 헤어의 방향성 이미지, 우측 : 데이터 셋에서 탐색된 유사한 헤어 방향성 이미지)

어스타일의 헤어 모델을 제공해줄 수 있을 뿐만 아니라 후보군의 헤어 모델을 추천하여 사용자가 원하는 헤어스타일을 손쉽게 아바타에 적용할 수 있다.

3. 3차원 얼굴 리깅 기술

복원된 아바타들은 메타버스에 구현된 3차원 가상 회의실에서 미팅을 하게 된다. 이런 가상환경에서 각각의 참석 인원들을 대표하는 아바타들은 발표를 하거나 논의를 하는 등 실제 미팅에서 하는 것과 똑같은 상호작용을 통해 정보교류를 하게 될 것이다. 이때 음성 정보 뿐만 아니라 상대방의 실시간 표정 변화도 그 의중을 파악하기 위한 중요한 단서로 작용한다. 따라서 실시간으로 참석자들의 표정을 재현하는 것이 필요하다.

- 얼굴 모델 기반 리깅 기술 : 표정의 재현은 일반적으로 생성된 3차원 얼굴모델을 리깅 하는 방식과 미리 설정된 모션을 이용하는 방식이 있다. 3차원 얼굴 모델을 리깅하는 방식은 3차원 얼굴 모델 복원 기술과 유사한 방식으로 생성된 3차원 얼굴모델 기



<그림 7> 기저벡터들의 가중치 변화에 따른 표정 재현

<그림 8> 기저벡터들의 가중치 변화에 따른 표정 재현 결과

<그림 9> 모션기반 표정재현

<그림 10> 언리얼 엔진을 통해 렌더링한 실사 복원 3D 아바타

저벡터들의 가중치를 변화시키는 방식으로 표정을 재현한다[12, 13].

- 모션 기반 리깅 기술 : 얼굴 모델을 직접 변형하는 것이 아닌 미리 설정된 모션을 이용하는 방식은 얼굴 모델의 대응되는 표정 변화에 따른 각 정점들의 대략적인 움직임을 사전에 정의하여 저장해 놓는다. 특정된 표정 변화는 특정된 랜드마크의 움직임으로 이해할 수 있는데, 웹캠을 통해 실시간으로 입력되는 사용자의 얼굴 영상에서 랜드마크들의 변화를 추적하고 움직임을 분석하여 대응되는 표정을 예측한 후 미리 정해 놓은 모션을 실행하면 자연스러운 표정을 재현할 수 있다. 본 과제에서는 AVATAR SDK를 이용하여 표정을 재현하였다[6].

Ⅲ. 맺음말

메타버스 세계는 기술의 진보와 함께 필연적으로 도래하게 될 미래라고 할 수 있다. 전 세계를 충격에 빠뜨린 코로나 바이러스의 대규모 유행은 관련 기술의 발전을 가속화하고 있다. 일상생활처럼 여겨지던 회의가 부담스러워지는 지금, 회의실을 메타버스 공간으로 옮겨가려는 시도는 좋은 해결책이라고 할 수 있다. 본 고에서는 이런 메타버스 공간에서 미팅을 진행할 때 실제 회의 상황과 비슷한 체험을 제공하기 위한 실감형 아바타의 생성 기술에 대해 설명을 하였다. 본 고에서 서술하고 있는 기술들은 현재 개발중인 기술로서, 머지않은 미래에 메타버스 공간에서 자신의 아바타를 이용하여 사실감 있는 미팅을 진행할 수 있을 것이다.

참 고 문 헌

[1] 메타버스, https://ko.wikipedia.org/wiki/%EB%A9%94%ED%83%80%EB%B2%84%EC%8A%A4

[2] Mirella Walker, Changing the Personality of a Face: Perceived Big Two and Big Five Personality Factors Modeled in Real Photographs, Journal of Personality and Social Psychology 110(4): 609-624

[3] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter, A 3D Face Model for Pose and Illumination Invariant Face Recognition, AVSS 2009

[4] Richard Hartley, Andrew Zisserman, Multiple View Geometry in computer vision

[5] 허재영, 장성걸, 박종일, 실감 영상회의 시스템을 위한 헤어스타일 탐색 방법, 한국방송미디어공학회 하계학술대회, 2021

[6] AVATAR SDK, https://avatarsdk.com/

[7] Xuehan Xiong, Fernando De la Torre, Supervised Descent Method and its Applications to Face Alignment, CVPR 2013

[8] Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski, Deep Alignment Network: A convolutional neural network for robust face alignment, CVPR 2017

[9] Xiaojie Guo, Siyuan Li, Jinke Yu, Jiawan Zhang, Jiayi Ma, Lin Ma, Wei Liu, Haibin Ling, PFLD: A Practical Facial Landmark Detector, arXiv preprint arXiv:1902.10859 (2019)

[10] Xinyao Wang, Liefeng Bo, Li Fuxin, Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression, ICCV 2019

[11] Li, Zuoxin, and Fuqiang Zhou, FSSD: feature fusion single shot multibox detector, arXiv preprint arXiv:1712.00960 (2017)

[12] P. Huber, Z. Feng, W. Christmas, J. Kittler, M. Rätsch, Fitting 3D Morphable Models using Local Features, ICIP 2015

[13] P. Huber, G. Hu, R. Tena, P. Mortazavian, W. Koppen, W. Christmas, M. Rätsch, J. Kittler, A Multiresolution 3D Morphable Face Model and Fitting, VISAPP 2016

[14] Jungsik Park, Byung-Kuk Seo, Jong-Il Park, A Framework for Real-Time 3D Freeform Manipulation of Facial Video, November 2019, Applied Sciences 9(21): 4707

[15] Zhou, Yi, et al, Hairnet: Single-view hair reconstruction using convolutional neural networks, ECCV 2018

[16] Zhang, Meng, and Youyi Zheng, Hair-GAN: Recovering 3D hair structure from a single image using generative adversarial networks, Visual Informatics 3.2 (2019): 102-112

[17] Doersch, Carl, Tutorial on variational autoencoders, arXiv preprint arXiv: 1606.05908 (2016)

[18] Bank, Dor, Noam Koenigstein, and Raja Giryes, Autoencoders, arXiv preprint arXiv: 2003.05991 (2020)

[19] Zhou, Yijun, and James Gregson, WHENet: Real-time Fine-Grained Estimation for Wide Range Head Pose, arXiv preprint arXiv:2005.10353 (2020)

[20] Bertók, Kornél, Levente Sajó, and Attila Fazekas, A robust head pose estimation method based on POSIT algorithm, Argumentum 7 (2011): 348-356

[21] 차세대 실감형 원격미팅 시스템 개발, 과학기술정보통신부 과제 No. 2020-0-0271

필 자 소 개

장 성 걸

- 2015년 : 한양대학교 컴퓨터소프트웨어학과 석사
- 2016년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학과 박사과정
- 2019년 ~ 현재 : ㈜아리아엣지 연구원
- 주관심분야 : 안면정보 분석 및 응용기술, 영상처리, 기계학습

김 형 민

- 2004년 : 백석대학교 컴퓨터공학과 학사
- 2017년 : 한양대학교 전자컴퓨터통신공학과 석사
- 2017년 ~ 현재 : 한양대학교 전자공학과 박사과정
- 2020년 ~ 현재 : ㈜아리아엣지 연구원
- 주관심분야 : 6D object pose estimation, 증강현실, 3차원 컴퓨터비전

서 병 국

- 2006년 : 한양대학교 전자공학과 학사
- 2008년 : 한양대학교 전자공학과 석사
- 2014년 : 한양대학교 전자공학과 박사
- 2014년 ~ 2016년 : Fraunhofer IGD Postdoctoral Research Fellow
- 2016년 ~ 현재 : ETRI 통신미디어연구소 차세대컨텐츠연구본부 CG/Vision Research Lab
- 주관심분야 : 6D pose estimation/localization and mapping, 3차원 컴퓨터비전, 증강현실

박 종 일

- 1987년 : 서울대학교 전자공학과 공학사
- 1989년 : 서울대학교 전자공학과 공학석사
- 1995년 : 서울대학교 전자공학과 공학박사
- 1992년 ~ 1994년 : 일본 NHK방송기술연구소 객원연구원
- 1995년 ~ 1996년 : 한국방송기술개발원 선임연구원
- 1996년 ~ 1999년 : 일본 ATR지능영상통신연구소 연구원
- 1999년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학부 교수
- 2018년 ~ 현재 : ㈜아리아엣지 대표이사
- 주관심분야 : 증강현실/가상현실, 3차원 컴퓨터비전, 인간컴퓨터상호작용