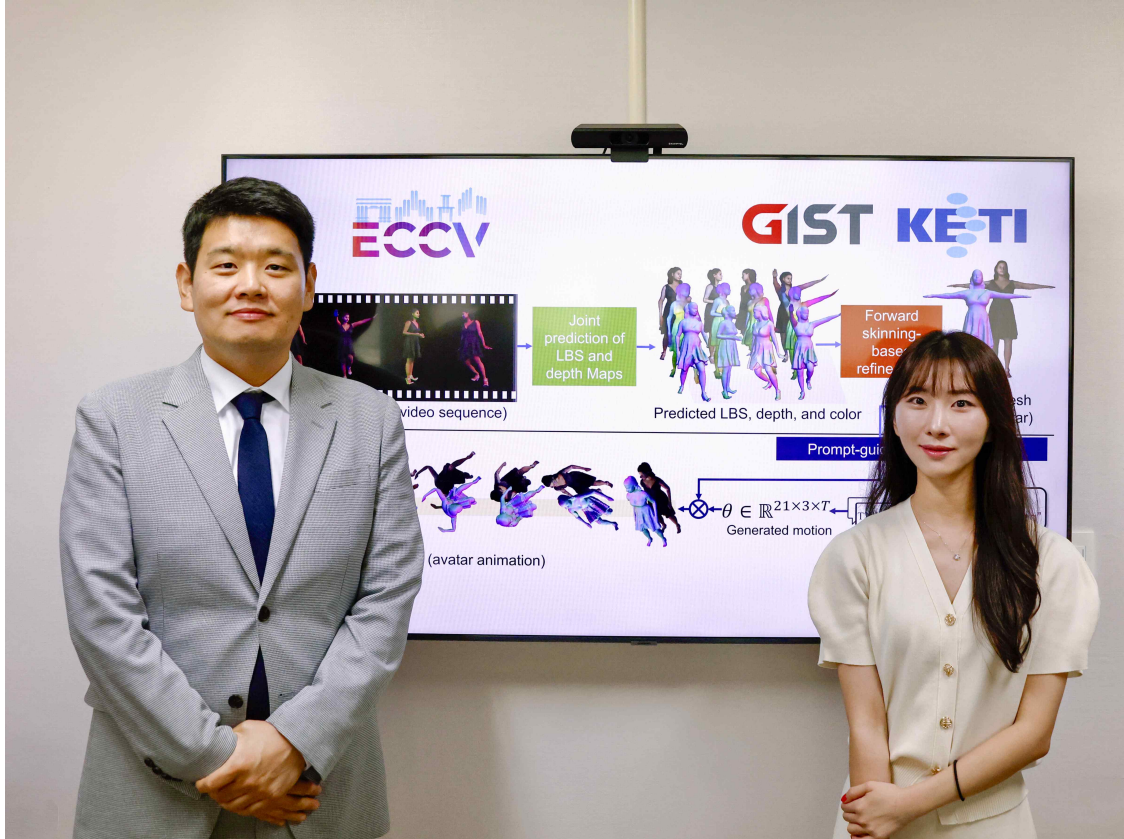


GIST, 2차원 평면 이미지만으로 움직이는 3차원 아바타 생성하는 AI 기술 개발

- AI대학원 전해곤 교수팀, 단순한 동작을 취하고 있는 단일 인물의 이미지를 별도의 복잡한 과정 없이도 움직임이 자연스러운 3차원 아바타로 구현하는 인공지능 알고리즘 개발
- 그래픽 전문가 손길을 거쳐야 하는 현재의 기법과 비교해 시간·비용 크게 절감... "인간이 수행하기 어려운 복잡한 동작 요하는 액션 영화·게임 등 문화산업에 활용 기대"
- 이탈리아 밀라노에서 열리는 세계 3대 컴퓨터 비전 학회 'ECCV 2024'에서 10월 2일 발표 예정



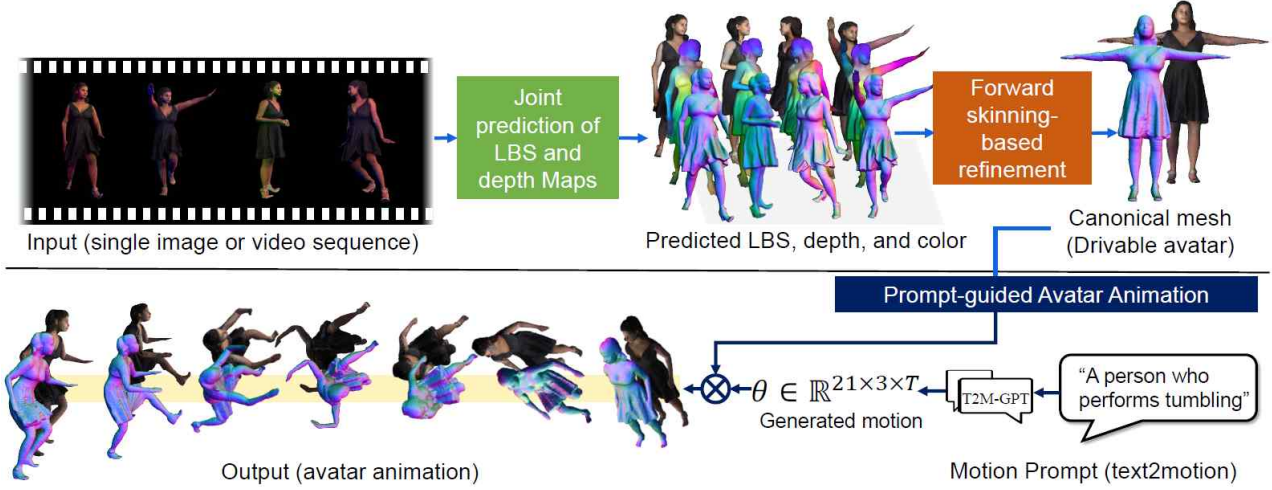
▲ (왼쪽부터) AI대학원 전해곤 교수, 신지수 석박통합과정생

2차원 이미지 혹은 영상만으로 어떤 동작이든 수행할 수 있는 고품질 3차원의 아바타를 누구나 만들 수 있는 인공지능(AI) 기술이 개발됐다. 영상 속에서 실제 인물과 같은 자연스러운 동작을 만드는 데 활용될 수 있을 것으로 기대된다.

광주과학기술원(GIST, 총장 임기철)은 AI대학원 전해곤 교수 연구팀이 **2차원 영상으로부터 다양한 동작이 가능한 고품질 3차원 인물을 구현하는 기술**을 개발했다고 밝혔다.

인공지능(AI) 기술을 통해 다중 시점의 카메라와 고가 장비 없이도 손쉽게 2차원 이미지만으로 3차원 인물 모델을 생성할 수 있는 혁신적인 방법이다.

또한 별도의 촬영 장비와 스튜디오 시설을 필요로 하지 않아 **비용과 자원을 크게 절감**할 수 있으며, 특히 게임, 영화, AR·VR 등 다양한 문화산업 분야에서 **실제 인물과 같은 사실적인 캐릭터를 생성하는 데 활용**될 수 있다.



▲ **본 연구의 요약 및 어플리케이션.** 단일 인물에 대한 이미지를 통해 각각의 3차원 모델을 구현하고, 해당 모델들의 정보를 통합하여 움직임이 가능한 사실적인 아바타를 만들어 낸다.

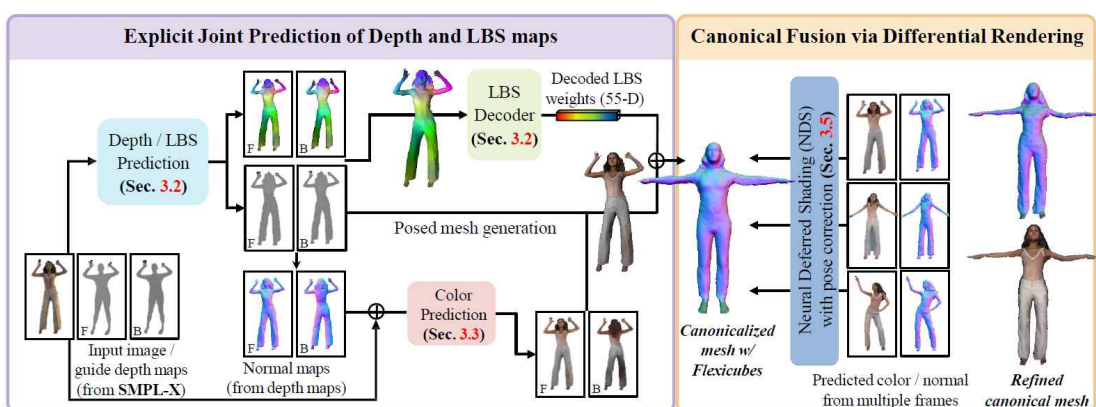
단일 이미지에서 3차원 인물을 생성하는 기존 방법론들은 정적인 아바타로, 단순히 3차원 공간에서 형태를 만드는 데 그치며 **자유로운 움직임이 불가능하다는 한계**를 지닌다.

이러한 정적 아바타를 실제로 움직이게 하려면 캐릭터에 가상의 뼈대와 관절을 설정하여 각 부위가 자연스럽게 움직일 수 있도록 만드는 **'리깅(rigging)'**이라는 과정이 필수적이다.

이 과정은 매우 정교한 작업을 요구하며, 캐릭터의 움직임이 현실감 있게 표현되기 위해서는 **숙련된 그래픽 디자이너들이 모든 관절과 움직임의 범위를 하나하나 수동으로 지정**해야 한다.

따라서 움직일 수 있는 캐릭터를 만들기 위해서는 여전히 상당한 전문 지식과 리소스가 필요하며, 최근에는 **자동화된 리깅 기술이나 보다 효율적인 애니메이션 기법에 대한 연구가** 진행되고 있다.

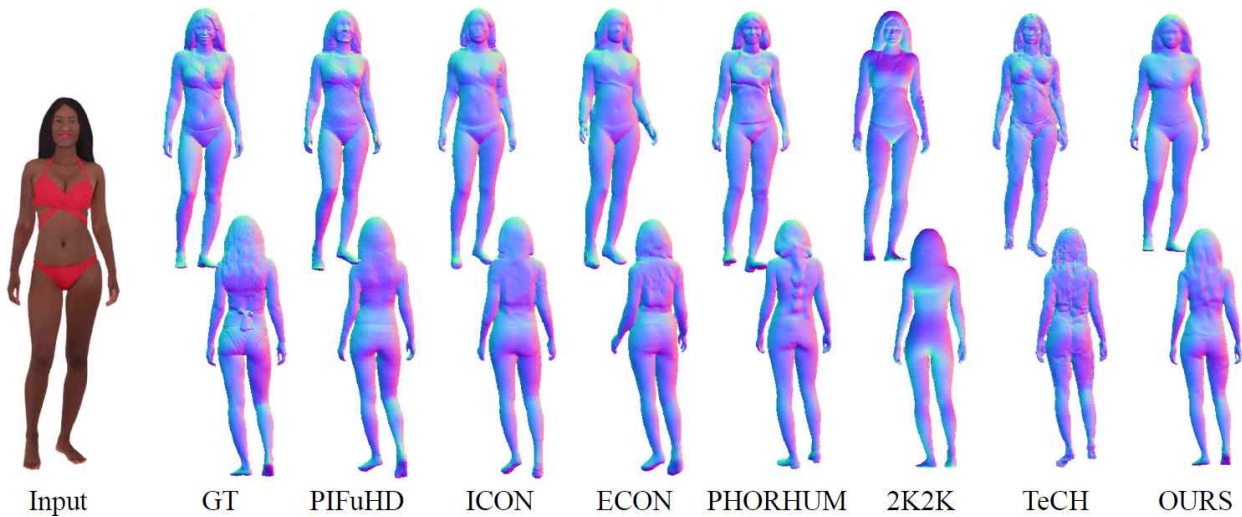
이번 연구의 핵심은 **깊이 정보와 스키닝 맵(Linear Blend Skinning, LBS Map)*을 동시에 예측**한 후, 이를 바탕으로 초기 다이내믹 아바타를 생성하고, 여러 이미지 정보를 통합 및 개선하는 것이다.



▲ **본 연구에서 제안한 아바타 생성 파이프라인.** 이미지에서 깊이 맵, 스키닝 맵, 컬러 맵을 뽑아서 3차원 메쉬를 만드는 부분과, 구현된 각 메시들을 통합하여 노멀맵과 컬러맵 차이 기반 개선을 통해 최종적인 아바타를 만드는 부분으로 나뉘어진다.

한 인물이 몇 가지 포즈를 취한 이미지만으로도 3차원 인물 모델의 생성이 가능한 이 기술은 먼저, 각 이미지에서 **제1신경망을 통해 인물의 앞뒤 깊이 정보와 스키닝 맵을 예측한다.** 높은 차원의 스키닝 맵을 간단히 표현하기 위해 **오토 인코더 및 디코더로 구성되어 있는 제2신경망 모델을 사용하여 복잡한 스키닝 정보를 압축하고, 다시 풀어내어 정확한 움직임**을 구현한다.

* **스키닝 맵(Linear Blend Skinning Map, LBS Map):** 본 자료에서 사용되는 스키닝 맵이란 3차원 인물의 각 관절 움직임 방식 정보를 인물 이미지의 각 픽셀마다 담고 있는 2차원 맵을 지칭한다.



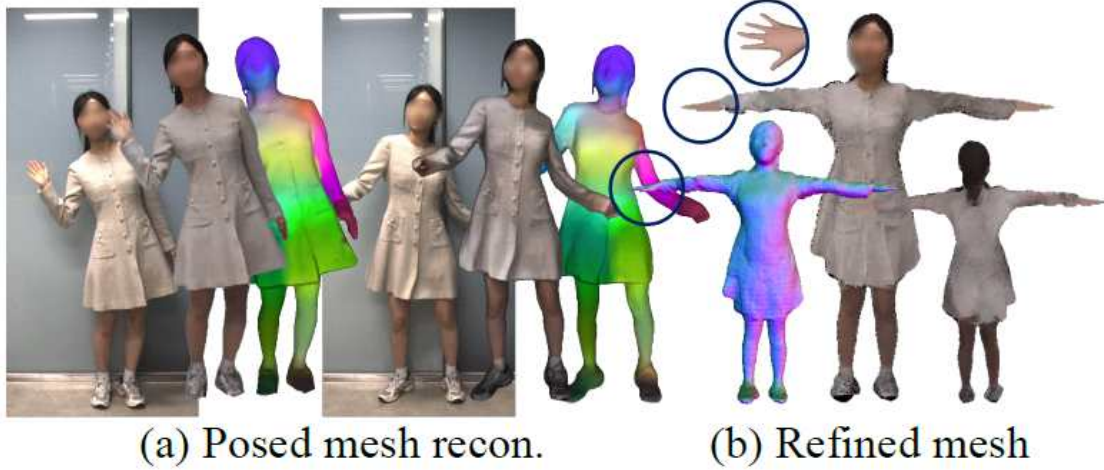
▲ **입력 이미지를 기반으로 3차원 구현을 수행한 기존 결과물과 본 연구진이 제안한 방식(Ours)으로 구현한 결과물의 앞뒤 노멀 맵 비교.**

그 다음, 예측된 깊이 정보와 스키닝 맵을 바탕으로 초기 3차원 아바타, 즉 **캐노니컬 메시***를 생성한다. 기존에는 인물 템플릿 모델을 활용하여 구현된 모델에 대해 스키닝 값을 계산하고 캐노니컬 메시를 생성해야 했으나, 연구팀이 개발한 기술을 이용하면 별도의 후처리 과정 없이도 캐노니컬 메시를 생성할 수 있다.

* **캐노니컬 메시(Canonical mesh):** 3차원 인물 모델의 기본 포즈. 이후 다양한 동작을 할 수 있는 출발점으로 사용되며, 보통 팔을 벌리고 가만히 서있는 포즈로 정의된다.

다음으로, 연구팀이 제안한 포워드 스키닝 기반의 렌더링 기법을 사용해 **초기 아바타를 정교하게 다듬고, 텍스처를 입힌다.**

이 기법은 여러 이미지의 정보를 통합해 **컬러와 위치 차이를 2차원 이미지 기반으로 줄여 가며 아바타의 정확도를 높일 수 있다.** 또한, 생성된 아바타의 관절과 제1신경망을 통해 생성된 아바타의 포즈 차이를 줄여 **포즈 오차가 크지 않다는** 장점이 있다. 이러한 방식을 통해 자연스러운 움직임이 가능한 3차원 아바타를 만들 수 있다.



(a) Posed mesh recon.

(b) Refined mesh

▲ 넉넉한 핏의 원피스를 입은 실제 사람을 촬영하여 본 알고리즘을 통해 생성된 아바타의 예시. 또한 이 기술은 입력 이미지의 수량에 제한을 받지 않으며, 단일 이미지로도 현실감 있는 결과물을 얻을 수 있다는 장점이 있다. 특히, 최근 주목받고 있는 초거대 언어 모델 기반의 텍스트로 움직임을 생성하는 기술과 결합하여 **아바타가 다양한 동작을 구현할 수 있다는** 가능성을 보여준다.



▲ '가볍게 춤을 추는 동작'이라는 텍스트를 제공하였을 때, 본 연구의 알고리즘을 통해 생성한 아바타가 춤을 추는 모습. 어플리케이션의 한 예.

전해곤 교수는 "이번 연구 성과를 통해 기존 그래픽 디자이너들이 3차원 객체를 움직이는 데 **소요되는 시간과 비용을 크게 절감**할 수 있으며, 일반인들도 2차원 이미지만으로 **손쉽게 3차원 아바타를 생성하고 활용**할 수 있다"면서 "향후 인간이 수행하기 어려운 복잡한 동작을 요구하는 액션 영화 혹은 게임 등 다양한 문화산업에서 **사실적인 아바타로서 활용될 것으로 기대된다**"고 말했다.

AI대학원 전해곤 교수 연구실의 신지수 석박통합과정생이 주도하여 GIST가 한국전자기술연구원(KETI)과 공동으로 수행한 이번 연구는 국제 R&D 협력 과제인 뉴럴 휴먼 모델링 기반의 사용자 참여형 메타버스 공연 솔루션 개발(산업통상자원부, 한국산업기술진흥원) 사업과 다중 소스 영상의 객체 미디어 처리 기술 개발(과학기술정보통신부, 정보통신기획평가원)의 지원을 받았으며, 컴퓨터 비전 분야 세계 3대 학술대회로 꼽히는 'ECCV(European Conference on Computer Vision)'에서 2024년 10월 2일 발표될 예정이다.

논문의 주요 정보

1. 논문명, 저자정보

- 저널명 : The 18th European Conference on Computer Vision 2024 (ECCV)
 - 한국정보과학회 및 BK21+ 기준 최우수 학술대회
- 논문명 : Canonical Fusion: Generating Drivable 3D Human Avatars from Multiple Images
- 저자 정보 : 신지수(제1저자, GIST AI대학원), 이준명(제2저자, GIST 전기전자컴퓨터공학부), 이성민(제3저자, GIST AI대학원), 박민규(제4저자, 한국전자기술연구원), 강주미(제5저자, 한국전자기술연구원), 윤주홍(제6저자, 한국전자기술연구원), 전해곤(교신저자, GIST AI대학원)