

3D 가상인간 기술 동향

□ 박병서 / 오모션 주식회사

1. 서론

3D 가상인간 기술의 핵심 방향은 단순히 ‘더 사실적인 하나의 생성 모델’을 만드는 것이 아니라, 다양한 기술을 하나의 통합 구조로 결합하는 데 있다. 구체적으로는 인간 신체 구조 기반 모델, 인공지능 렌더링, 음성 및 표정 애니메이션, 행동 계획, 멀티모달 인지, 물리 시뮬레이션, 그리고 신뢰 및 규제 대응 기능까지 모두 통합하는 것이 중요하다. 최근 연구에서는 3D 인간 아바타 기술을 두 가지 축으로 구분한다. 첫째는 실제 데이터를 기반으로 인간을 복원하는 ‘재구성(reconstruction)’이며, 둘째는 새로운 인간 모델을 만들어내는 ‘생성(generation)’이다. 또한 상호작용 기술은 음성, 시선, 표정, 자세, 멀티모달 처리 등으로 나뉘어 각각 발전하고 있으며, 이는 3D 가상인간이 단순한 그래픽 객체가 아니라 실제처럼 반응하는 상호작용 시스템으로 진화하고 있음을 의미한다[1].

현재 기술 수준은 이미 연구 단계를 넘어 상용화 가능한 수준에 도달하였다. 예를 들어, Epic의 MetaHuman은 고품질 디지털 인간 생성 환경을 제공하며, NVIDIA ACE는 음성·AI·애니메이션·그래픽을 통합한 구조를 제시한다.

또한 Audio2Face-3D는 음성과 감정을 기반으로 얼굴 애니메이션을 실시간에 가깝게 생성하고 있다. 학계에서도 기술 발전이 빠르게 이루어지고 있다. InstantAvatar, GauHuman, SplattingAvatar 등의 연구는 단일 카메라 기반 3D 재구성과 실시간 렌더링을 가능하게 했으며, 짧은 학습 시간과 높은 처리 속도를 동시에 달성하고 있다. 중장기 기술 발전 방향은 크게 네 가지로 정리된다. 첫째, SMPL 및 FLAME과 같은 인체 기본 모델과 NeRF, Gaussian Splatting 같은 고품질 렌더링 기술의 결합이다. 둘째, LLM 및 VLM 기반 멀티모달 인지 기술을 통해 가상인간의 지능과 행동 능력을 강화하는 것이다. 셋째, 동작의 의미를 유지하는 리타겟팅 및 행동 생성 기술이 중요해지고 있다. 넷째, 품질 평가와 신뢰성 확보가 핵심 경쟁 요소로 부상하고 있다[2].

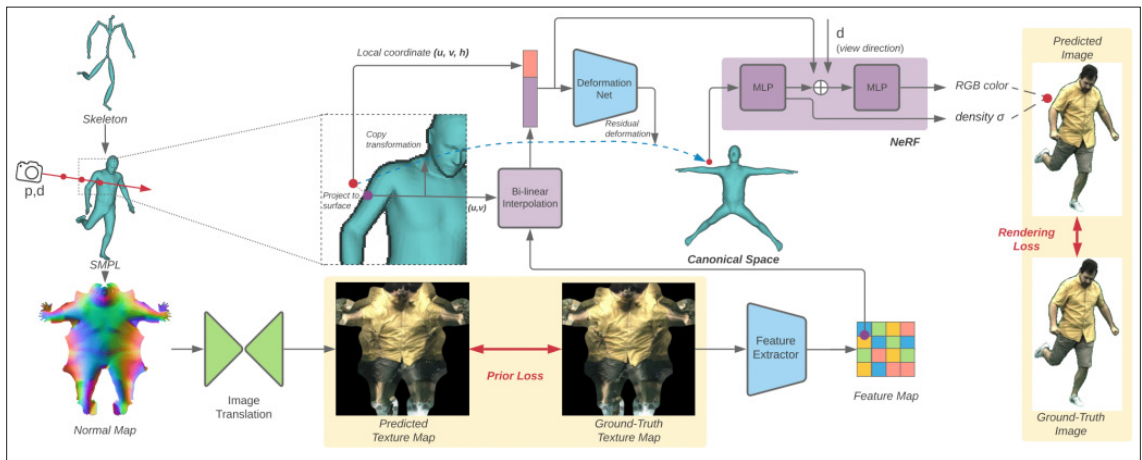
최근 연구에서는 3D 가상인간을 단순한 3D 캐릭터가 아니라, 실제 신체를 가진 지능형 대화 에이전트로 정의하는 경향이 강하다. 즉, 음성·텍스트·표정·시선·제스처 등 다양한 채널을 통해 사람과 상호작용하는 ‘체화된 시스템(embodied system)’으로 이해된다. 특히 2024년 기준 연구에서는 상호작용 요소를 음성, 얼굴 표정, 시선, 자세, 멀티

모달 인터랙션 등으로 나누어 분석하고 있으며, 3D 아바타 기술 또한 ‘재구성(reconstruction)’과 ‘생성(generation)’을 동시에 다루는 방향으로 발전하고 있다. 이는 ‘보이는 외형(몸)’과 ‘행동하는 지능’이 분리된 문제가 아니라, 하나의 통합된 시스템이라는 점을 의미한다[3]. 본 보고서에서는 3D 가상인간 기술을 총 여섯 가지 평가 기준으로 분석한다. 첫째, 사실성은 외형, 광학 표현, 음성의 자연스러움을 의미한다. 둘째, 통제 가능성은 페르소나(identity), 자세(pose), 표정(expression), 감정(emotion), 제스처 등을 얼마나 정밀하게 제어할 수 있는지를 의미한다. 셋째, 실시간성은 학습 시간, 추론 속도, 프레임 속도(FPS), 그리고 사용자 입력에 대한 반응 지연 등을 기준으로 판단한다. 넷째, 상호작용성은 대화 흐름(turn-taking), 상황 인식, 도구 사용, 기억 기능, 사회적 신호 처리 능력 등을 포함한다. 다섯째, 운영 상용성은 브라우저, 모바일, 클라우드, 게임 엔진 등 다양한 환경과의 연동 가능성을 의미한다. 여섯째, 신뢰성은 개인정보 보호, 저작권, 딥페이크 식별, 안전 정책, 품질 평가 체계 등을 포함하는 개념이다[4]. 특히 최근에는 표준과 정책의 발전으로 인해 ‘신뢰성’ 요소의 중요성이 크게 증가하고 있으며, 이는 기술 성능뿐 아니라 서비스 운영 측면에서도 필수 요소로 자리잡고 있다.

3D 가상인간 기술의 시작은 1989년 Visible Human

Project로 거슬러 올라간다. 이후 기술은 크게 네 단계로 발전하였다. 초기에는 수작업 기반 3D 자산 제작이 중심이었고, 이후 파라메트릭 인체 및 얼굴 모델이 등장하였다. 그 다음 단계에서는 신경 렌더링 기반 아바타 기술이 발전하였으며, 최근에는 멀티모달 기반 인공지능 에이전트 형태로 진화하고 있다. 2014년 전후 시기는 ‘데이터와 사전 모델(prior)’이 중요한 역할을 하던 시기였다. Human3.6M 데이터셋은 대규모 3D 인체 자세 데이터를 제공하여 연구의 기준이 되었고, GAN은 사실적인 데이터 생성을 위한 핵심 구조를 제시하였다. 이어 SMPL 모델은 인체를 수학적으로 표현할 수 있는 표준 구조를 제공하였다. 이 모델은 다양한 체형과 자세를 표현할 수 있으며, 기존 그래픽 시스템과도 호환되어 이후 3D 인체 기술의 기본 좌표계로 자리잡았다[5]. 2017년부터 2020년 사이에는 얼굴, 음성 표현 기술이 크게 발전하였다. Transformer 구조는 이후 LLM과 VLM의 기반이 되었고, FLAME 모델은 정교한 얼굴 표현을 가능하게 했다. 또한 Tacotron 2는 자연스러운 음성 합성을 구현하였고, NeRF는 3D 장면을 신경망으로 표현하는 새로운 방법을 제시하였다. 2021년 이후에는 ‘인간 특화 신경 아바타’ 기술이 본격적으로 발전하였다[6].

Neural Actor와 IMavatar는 실제 사람과 유사한 3D 인간 표현을 가능하게 했으며, InstantAvatar는 매우 짧



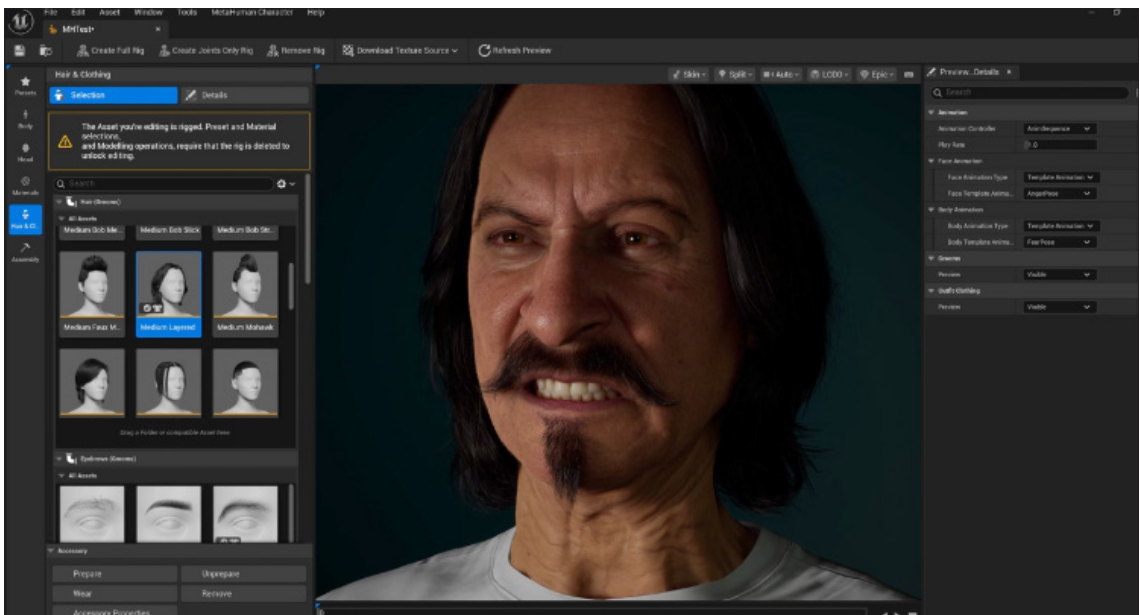
<그림 1> Neural Actor 알고리즘

은 시간 내 학습과 실시간 렌더링을 구현하였다. 또한 GauHuman과 SplattingAvatar는 매우 높은 프레임 속도를 달성하여 실시간 처리 기술의 중요한 전환점을 만들었다. 한편 생성 기술 측면에서는 EG3D와 DreamAvatar가 등장하여, 텍스트나 조건을 기반으로 3D 인간을 생성하는 기술이 발전하였다. 이로 인해 ‘재구성’과 ‘생성’의 경계가 점점 흐려지고 있다. 다만 실제 제품 관점에서는 완전히 자유로운 생성보다는, 기존 모델과 제어 변수를 활용한 ‘제어 가능한 생성 방식’이 더 현실적인 방향으로 평가된다[7].

II. 3D 가상인간의 핵심 기술

현재 3D 가상인간 시스템은 단일 기술이 아니라 여러 계층이 결합된 구조로 설계되는 것이 가장 일반적이다. 기본적으로 인체와 얼굴의 구조를 정의하는 모델이 기반이 되고, 그 위에 음성, 표정, 동작 생성 기능이 실시간으로 작동하며, 최상위에는 대화와 행동을 결정하는 인공지

능 계층이 위치한다. 이 구조는 상용 시스템에서도 공통적으로 나타난다. MetaHuman, NVIDIA ACE, UneeQ, DeepBrain AI 등은 표현 방식은 다르지만, 모두 ‘신체 표현-상호작용-인공지능 제어’로 구성된 계층형 구조를 따른다. 전체 시스템 흐름은 다음과 같이 이해할 수 있다. 먼저 사용자의 입력(텍스트, 음성, 영상 등)이 들어오면, 이를 인식하는 지각 계층이 작동한다. 이후 대화 맥락과 사용자 상태를 추적하는 상태 관리 계층이 정보를 정리한다. 그 다음 단계에서는 LLM 또는 VLM 기반 인공지능이 상황을 이해하고, 어떤 행동을 할지 결정한다. 이 과정에서 발화 내용, 감정, 시선, 제스처, 동작 등이 함께 계획된다. 행동 계획 결과는 실제 표현 계층으로 전달된다. 여기서 음성 생성(TTS), 얼굴 애니메이션(립싱크 및 표정), 신체 동작 생성(IK, 모션 합성)이 각각 수행된다. 이후 SMPL, Mesh, FLAME, NeRF, Gaussian Splatting 등으로 구성된 표현 모델을 통해 3D 형태가 생성되며, 최종적으로 물리 시뮬레이션과 렌더링 과정을 거쳐 사용자에게 출력된다. 이 구조의 핵심 특징은 두 가지이다.



<그림 2> Unreal, MetaHuman

첫째, ‘행동 계획’과 ‘표현 생성’이 분리되어 있다는 점이다. 둘째, 인간 구조 모델과 신경망 기반 표현 기술이 함께 사용되는 하이브리드 방식이라는 점이다. 이러한 구조는 현재 3D 가상인간 시스템에서 가장 안정적이고 실용적인 설계 방식으로 평가된다.

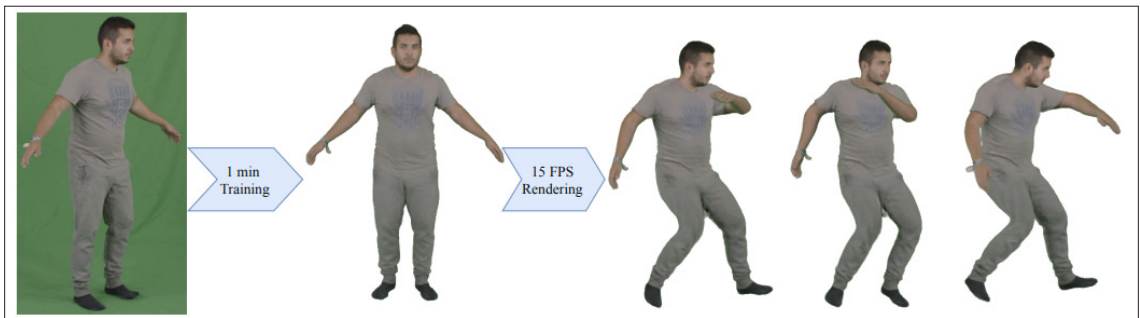
3D 가상인간 기술은 여러 핵심 구성 요소로 이루어지며, 각 기술이 결합되어 전체 시스템의 품질을 결정한다. 먼저, 신체 모델링 기술에서는 SMPL과 같은 파라메트릭 인체 모델이 핵심 역할을 한다. 이 모델은 다양한 체형과 자세를 수학적으로 표현할 수 있으며, 기존 그래픽 엔진과의 호환성이 높아 기본 구조로 널리 사용된다. 얼굴 모델링에서는 FLAME 모델이 대표적이다. 이 모델은 얼굴 형태, 턱 움직임, 목 회전, 눈동자 움직임 등을 정밀하게 제어할 수 있어 자연스러운 표정 표현이 가능하다. 이러한 모델들은 최종 결과를 만드는 기술이 아니라, ‘제어 가능한 기준 좌표계’ 역할을 하며 다른 AI 기술과 결합될 때 큰 효과를 발휘한다. 다음으로 애니메이션 기술에서는 inverse kinematics(IK), 리타겟팅, 모션 합성이 중요하다. IK는 특정 위치에 손이나 발을 정확히 맞추는 기술이며, 실시간 환경에서도 빠르게 동작해야 한다.

최근에는 단순히 움직임을 맞추는 것을 넘어서, 행동의 의미까지 유지하는 ‘의미 기반 모션 리타겟팅’ 기술이 중요해지고 있다. 얼굴 애니메이션과 립싱크 기술은 두 가지 방식으로 발전하고 있다. 첫째는 카메라 기반으로 실제 얼굴 움직임을 캡처하는 방식이며, 둘째는 음성을 입력으로

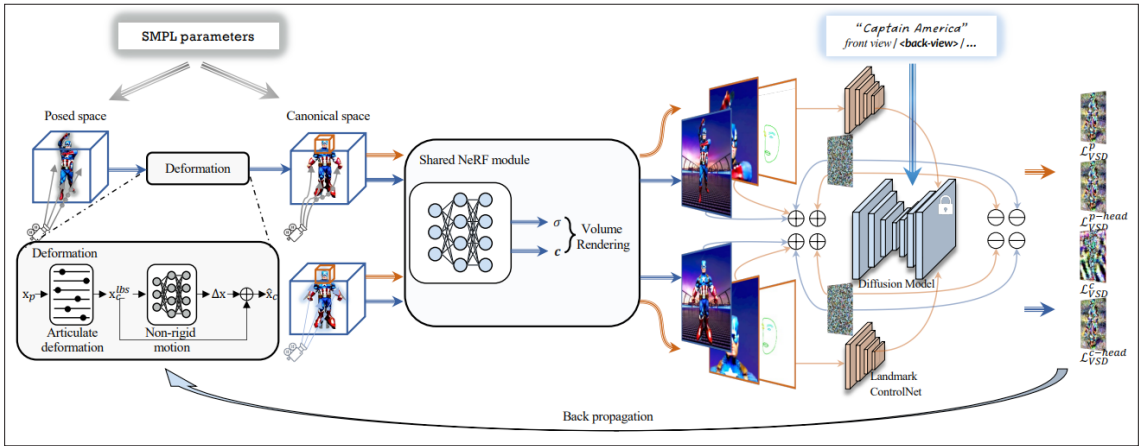
얼굴 움직임을 생성하는 방식이다. 실무에서는 이 두 방식이 경쟁 관계가 아니라, 고품질 제작에는 캡처 방식, 실시간 서비스에는 음성 기반 방식이 더 적합한 보완적 관계로 사용된다. 음성 생성 기술은 Tacotron 2와 VITS를 통해 크게 발전하였다. 현재는 음성 품질 자체보다 감정 표현, 반응 속도, 대화 흐름 제어가 더 중요한 요소가 되고 있다. 또한 실시간 상호작용에서는 speech-to-speech, 대화 중단 처리, 자연스러운 반응 속도가 핵심 경쟁 요소로 작용한다. 렌더링 기술에서는 NeRF가 중요한 전환점을 만들었다. 이 기술은 다양한 시점에서도 일관된 3D 표현을 가능하게 하지만, 초기에는 속도와 비용 문제가 있었다. 이를 개선하기 위해 InstantAvatar, GauHuman, SplatingAvatar 등의 기술이 등장하였으며, 빠른 학습과 실시간 렌더링을 동시에 달성하는 방향으로 발전하고 있다. 최근에는 순수 신경 렌더링 방식에서 벗어나, 메쉬 기반 구조와 결합한 하이브리드 방식이 주류로 자리잡고 있다. 마지막으로 물리 시뮬레이션 기술도 중요성이 증가하고 있다. 머리카락, 옷, 피부 움직임 등은 단순한 장식 요소가 아니라, 현실감을 결정하는 핵심 요소이다.

III. 인공지능과 3D 가상인간 기술의 결합

3D 가상인간 생성 기술에서는 GAN과 Diffusion 모



<그림 3> InstantAvatar: Learning Avatars from Monocular Video in 60 Seconds



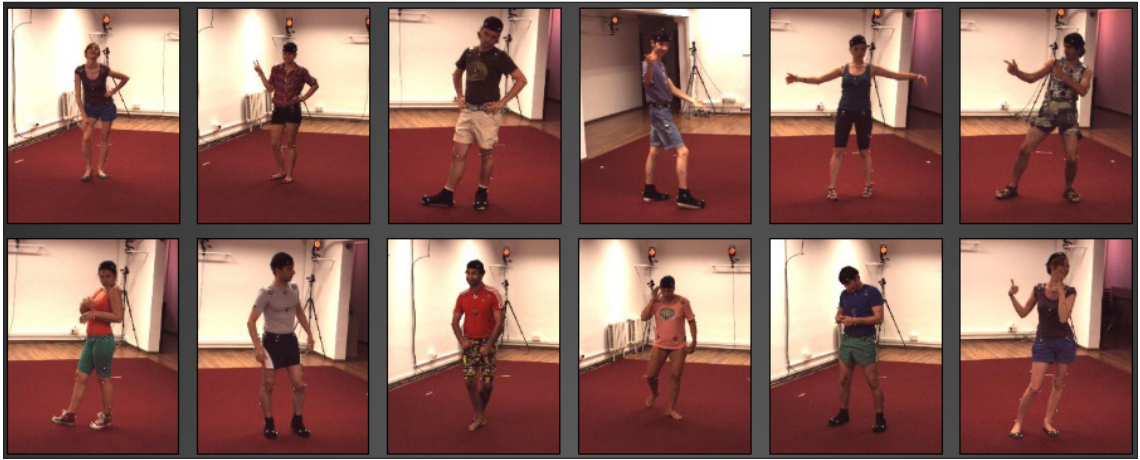
<그림 4> DreamAvatar, 신경망 구조

델이 서로 다른 역할을 하며 함께 발전하고 있다. 먼저, GAN(생성적 적대 신경망)은 빠른 생성 속도와 높은 효율성을 강점으로 가진다. EG3D와 같은 모델은 3D 구조를 고려한 이미지 생성이 가능하며, 여러 시점에서 관련된 결과를 제공할 수 있다. 반면, Diffusion 모델은 텍스트나 이미지 조건을 기반으로 더 정밀한 생성이 가능하다. DreamAvatar는 텍스트와 형태 정보를 동시에 활용하여, 자세와 형태를 제어할 수 있는 3D 인간 생성 기술을 제시하였다. 하지만 인간 모델은 다른 객체보다 구조가 훨씬 복잡하다. 신체 구조, 얼굴 특징, 의상, 머리카락, 행동 의미 등 다양한 요소가 동시에 고려되어야 하기 때문이다. 이 때문에 실제 시스템에서는 단순한 생성 모델만으로는 한계가 있다. 대신 CLIP, Diffusion, SMPL과 같은 사전 모델(prior)을 결합한 방식이 더 안정적이고 현실적인 접근으로 평가된다. 즉, 3D 가상인간 생성 기술은 ‘완전 자유 생성’보다는 제어 가능한 조건을 기반으로 하는 ‘제약 기반 생성(constraint-aware generation)’ 방향으로 발전하고 있다. 이러한 흐름은 실제 제품 개발에서도 중요하다. 완전히 새로운 인간을 만드는 것보다, 기존 구조를 유지하면서 원하는 형태와 동작을 정확하게 제어하는 것이 더 실용적이기 때문이다. 결론적으로, 생성 모델은 단독으로 사용되는 기술이 아니라 기존 인체 모델, 렌더링 기술, 애니메이션 시스템과 결합될 때 실제 서비스 수준의 품질을

구현할 수 있다.

IV. 데이터셋 및 평가지표

3D 가상인간 구현 및 인공지능 훈련을 위한 대표적인 데이터셋도 다양한 분야에 걸쳐 존재한다. Human3.6M은 3D 인체 자세 인식의 기준이 되는 데이터셋이며, AMASS는 다양한 모션 데이터를 통합하여 학습에 활용된다. 3DPW는 실제 환경에서의 3D 자세 데이터를 제공하여 현실 적용 성능을 검증하는 데 사용되며, VoxCeleb은 음성과 화자 인식 연구에 활용되는 대표 데이터셋이다. 또한 FLAME 기반 데이터는 얼굴 표현 모델링에 활용되며, DHHQA는 디지털 인간의 품질을 평가하기 위한 기준 데이터셋으로 사용된다. 하지만 현재 가장 큰 문제는 데이터가 부족한 것이 아니라, 각 데이터셋이 서로 다른 영역(신체, 얼굴, 음성, 상호작용 등)에 분리되어 있다는 점이다. 이로 인해 전체 시스템을 평가할 수 있는 통합적인 데이터셋이나 벤치마크는 아직 부족한 상황이다. 평가 지표 또한 여러 층위로 나뉜다. 신체 자세 정확도는 MPJPE 등의 지표로 평가하며, 렌더링 품질은 PSNR, SSIM, LPIPS 등을 사용한다. 실시간성은 FPS와 지연 시간으로 평가되고, 음성 품질은 MOS와 화자 유사도 등을 기준으로 판단한다.



<그림 5> Human3.6M, 데이터셋

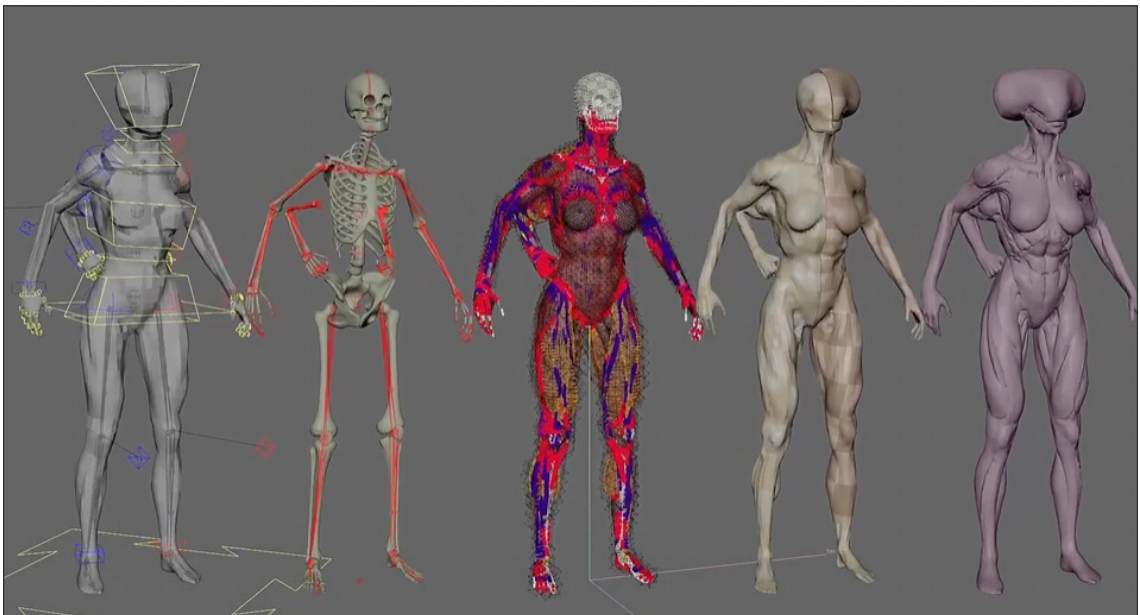
또한 립싱크 정확도, 사용자 만족도, 상호작용 성공률 등도 중요한 평가 요소로 활용된다.

특히 중요한 점은, 단일 지표만으로 전체 품질을 평가할 수 없다는 것이다. 예를 들어, 영상 품질이 높아도 표정이나 타이밍이 어색하면 사용자 경험은 낮아질 수 있다. 따라서 3D 가상인간 기술은 다양한 평가 지표를 함께 사용

하는 ‘다차원 평가 방식’이 필수적이다.

V. 물리 시뮬레이션 및 사실성 요소

물리 시뮬레이션은 3D 가상인간의 현실감을 결정하는



<그림 6> Ziva, 근육, 지방, 피부 시뮬레이션

핵심 기술로 중요성이 크게 증가하고 있다. 특히 머리카락, 의상, 피부, 근육과 같은 요소는 단순한 시각 효과와 아니라, 사람이 느끼는 ‘자연스러움’에 직접적인 영향을 준다. 예를 들어 Unity의 Enemies 데모는 고급 피부 표현, 머리카락 처리, 그리고 실시간 변형 기술을 통해 매우 높은 수준의 디지털 인간을 구현하였다. 또한 Ziva 기술은 근육, 지방, 피부의 움직임을 물리적으로 시뮬레이션하여 보다 현실적인 인체 변형을 가능하게 한다. PhysX와 같은 물리 엔진은 의상(Cloth), 연체 물질(Soft Body), 입자(Particle), 유체(Fluid) 등 다양한 물리 현상을 처리할 수 있으며, 캐릭터 움직임까지 포함하는 종합적인 시뮬레이션 환경을 제공한다. 앞으로는 이러한 요소들이 선택이 아니라 필수 요소로 자리잡을 것으로 예상된다. 특히 머리카락, 의상, 피부와 같은 요소는 가상인간의 ‘지각적 현실감(perceptual realism)’을 결정하는 핵심 층으로 작용한다. 즉, 단순히 얼굴을 사실적으로 만드는 것만으로는 부족하며, 전체 신체의 물리적 움직임과 미세한 변화까지 자연스럽게 표현해야 진정한 현실감을 구현할 수 있다.

VI. 3D 가상인간 산업 및 생태계

현재 3D 가상인간 산업은 세 가지 주요 계층으로 나뉘어 발전하고 있다. 첫째, 캐릭터 생성 계층(character creation layer)이다. 이 영역은 MetaHuman, Unity/Ziva와 같이 모델링, 리깅, 텍스처, 헤어, 물리 시뮬레이션 등을 담당한다. 둘째, 상호작용 인공지능 계층(interaction brain layer)이다. Inworld, Convai와 같은 기업들이 해당하며, 언어 이해, 행동 결정, 상황 인식 등 AI 기반 인터랙션 기능을 제공한다. 셋째, 실행 및 운영 계층(runtime/orchestration layer)이다. NVIDIA ACE, UneeQ, Soul Machines, DeepBrain AI 등이 여기에 해당하며, 음성, 애니메이션, 클라우드 배포 등을 통합하여 실제 서비스를 구현한다. 이처럼 산업은 하나의 기업이 모든 기술을 담당하기보다, 각 계층을 나누고 서로 결합하는 형태로 발전하고 있다. 주요 기업별 특징을 보면 다음과 같다. MetaHuman은 고품질 캐릭터 제작과 리깅 기능에 강점이 있으며, NVIDIA ACE는 음성, AI, 애니메이션을 모뮬화하여 통합 제공하는 구조를 갖는다. Unity와 Ziva



<그림 7> 오모션의 초실사 가상인간 및 상호작용 기술

는 물리 기반 표현과 실시간 렌더링에서 강점을 가지며, Soul Machines는 기업용 디지털 인간 운영 및 자동화에 초점을 둔다. Inworld는 실시간 음성 기반 AI 인터랙션에 강점을 가지며, UneqQ는 웹과 모바일 기반 서비스에 적합한 구조를 제공한다. Convai는 멀티모달 인지와 행동 중심 캐릭터 구현에 집중하고 있다. 국내 시장에서도 발전이 이루어지고 있다. 오모션 주식회사는 얼굴, 음성, 제스처를 결합한 실시간 디지털 휴먼 기술을 제공한다. DeepBrain AI는 얼굴과 음성, 제스처를 정밀하게 복제하는 AI Human 기술을 제공하며, 온마인드는 실시간 상호작용 기반 가상인간을 목표로 하고 있다. 중요한 점은 대부분의 상용 시스템이 ‘하이브리드 구조’를 사용한다는 것이다. 즉, 전통적인 그래픽 기술과 인공지능 기술을 결합하여 전체 시스템을 구성하고 있다. 이는 연구 분야에서 나타나는 기술 통합 흐름과 동일한 방향이며, 향후에도 플랫폼 중심 구조가 계속 강화될 것으로 예상된다.

Ⅶ. 윤리·법제·사회적 이슈

3D 가상인간 기술은 단순한 기술 문제가 아니라, 윤리·법·사회적 문제와 밀접하게 연결되어 있다. 첫째, 투명성 문제가 중요하다. EU AI Act는 AI로 생성된 이미지, 음성, 영상에 대해 인공지능 생성 여부를 명확히 표시하도록 요구하고 있다. 이는 가상인간이 실제 사람처럼 보일수록, 사용자에게 AI임을 명확히 알리는 기능이 필수라는 의미이다. 한국에서도 생성형 AI 결과물에 워터마크를 표시하는 정책이 추진되고 있으며, 이는 기술이 아닌 ‘서비스 설계 요소’로 고려되어야 한다. 둘째, 개인정보 및 생체정보 보호 문제이다. 얼굴, 목소리, 표정, 시선, 대화 기록 등은 모두 민감한 데이터에 해당한다. 따라서 데이터 수집 동의, 사용 목적 제한, 보관 기간, 제3자 제공 등의 정책을 초기 설계 단계에서 포함해야 한다. 특히 실제 사람을 기반으로 한 디지털 인간일수록, 이 문제는 단순한 기술 이슈가 아니라 제품 요구사항 수준으로 중요해진

다. 셋째, 저작권 문제이다. 3D 가상인간에서는 학습 데이터뿐 아니라, 특정 인물의 외형, 목소리, 행동 스타일까지 권리 문제가 발생할 수 있다. 또한 생성된 결과물의 소유권과 책임 문제도 함께 고려해야 한다. 따라서 데이터 출처(provenance)와 사용 동의 정보는 시스템 구조의 기본 요소로 포함되어야 한다. 넷째, 사회적 영향과 신뢰 문제이다. 가상인간은 실제 사람과 유사한 외형과 행동을 가지기 때문에, 사용자에게 강한 몰입감과 감정적 반응을 유도할 수 있다. 이는 교육, 상담, 고객 서비스 등에서는 장점이 될 수 있지만, 과도한 친밀감 형성이나 설득력 문제 등 부작용도 발생할 수 있다. 따라서 단순히 ‘얼마나 사람처럼 보이는가’가 아니라, ‘사용자와 어떤 관계를 형성하는가’까지 고려한 설계가 필요하다.

Ⅷ. 결론

향후 3D 가상인간 기술 개발은 하나의 뛰어난 모델을 만드는 것이 아니라, 전체 시스템을 어떤 구조로 통합하고 발전시킬 것인가가 핵심 과제가 된다. 기술 발전 우선 순위는 일반적으로 다음과 같은 단계로 정리된다. 먼저 데이터와 인체 모델(prior)을 정비하고, 그 다음 하이브리드 아바타 엔진을 구축한다. 이후 저지연 상호작용을 구현하고, 멀티모달 AI 에이전트로 발전시키며, 마지막으로 평가 기준과 규제를 시스템에 내재화하는 방향으로 진행된다. 이러한 로드맵은 현재 기술 성숙도와 실제 개발 과정에서 발생하는 병목을 반영한 것으로, 단순한 이론이 아니라 실무 중심의 개발 방향을 제시한다. 구체적인 핵심 개발 방향은 다음과 같이 정리할 수 있다. 첫째, 하이브리드 구조를 기본으로 해야 한다. SMPL과 FLAME 같은 인체 모델과 신경 렌더링 기술을 결합해야 실제 서비스 수준의 제어성과 품질을 확보할 수 있다. 둘째, 단순한 움직임이 아니라 ‘행동의 의미’를 유지하는 모션 기술이 중요하다. 즉, 정확한 자세보다 자연스럽고 이해 가능한 행동을 생성하는 것이 핵심이다. 셋째, 음성, 얼굴, 언어를 하나의

상호작용 루프로 통합해야 한다. 향후 경쟁력은 자연스러운 반응 속도와 감정 표현, 대화 흐름 처리 능력에서 결정된다. 넷째, 물리 시뮬레이션과 외형 표현 기술에 대한 투자가 필요하다. 특히 머리카락, 의상, 피부와 같은 요소는 전체 현실감을 크게 좌우한다. 다섯째, 전체 시스템을 평

가할 수 있는 통합 벤치마크 구축이 필요하다. 현재는 개별 요소 중심 평가가 많아, 전체 성능을 판단하기 어렵다. 여섯째, 신뢰성과 규제 대응을 기술 요소가 아닌 필수 요구사항으로 포함해야 한다. 이는 향후 서비스 경쟁력에 중요한 요소가 될 것이다.

참 고 문 헌

- [1] Wang, R., Cao, Y., Han, K., & Wong, K.-Y. K., "A Survey on 3D Human Avatar Modeling: From Reconstruction to Generation," *arXiv* (2024).
- [2] <https://www.metahuman.com/>
- [3] Wang, X., Cao, N., Chen, Q., & Cao, S., "The interaction design of 3D virtual humans: A survey," *Computer Science Review*, 53 (2024), pp. 100653.
- [4] <https://webstore.ansi.org/standards/ieee/ieee30792023>
- [5] Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C., "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7) (2014), pp. 1325-1339.
- [6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 30 (2017), pp. 5998-6008.
- [7] Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., De Mello, S., Gallo, O., Guibas, L., Tremblay, J., Khamis, S., Karras, T., & Wetzstein, G., "Efficient Geometry-aware 3D Generative Adversarial Networks," *arXiv* (2021).

저 자 소 개



박 병 서

- 2019년 : 광운대학교 경영학과 학사
- 2024년 : 광운대학교 전자재료공학과 박사
- 2025년 : 주식회사 한터글로벌, 서비스개발 본부장
- 2026년 ~ 현재 : 오모션 주식회사, CAIO/연구소장
- 주관심분야 : 3D 그래픽스, 인공지능