

디지털 휴먼 기술의 자세 생성 및 리타게팅 기술 현황

□ 조경민, 강현석, 이정우, 박구만 / 서울과학기술대학교

요약

본 논고는 디지털 휴먼 제작에 핵심적인 자세 생성과 모션 리타게팅 기술의 발전 과정 및 최신 동향을 분석한다. 과거 수학적 연산과 수작업에 의존하던 기술적 한계를 넘어, 최근 딥러닝과 디퓨전 모델의 도입을 통한 기술적 발전을 이끌어낸 핵심 논문들을 소개하며 두 기술들의 발전 양상을 조명한다. 이러한 생성형 AI를 통해 미래 산업 및 1인 콘텐츠 산업에 긍정적인 영향을 끼칠 수 있음을 예측하면서, 동시에 기술의 발전으로 인해 발생할 수 있는 사회적 문제점들을 우려하며 가이드라인의 필요성을 주장한다.

I. 서론

최근 온라인과 방송미디어에서 사람과 유사한 가상의 존재들을 쉽게 찾아볼 수 있다. <그림 1>에서 볼 수 있듯이 소셜 미디어에서 수백만 명의 팔로워와 소통하는 가상 인플루언서부터, 날씨예보와 그날의 소식을 알려주는 가상인간 아나운서까지 디지털 휴먼은 일상 곳곳에서 찾아볼 수 있다.

그러나 이러한 디지털 휴먼은 불과 몇 년 전까지만 해도 우리들에게 불쾌한 골짜기를 경험하게 했다. 1996년 일본에서 CG를 통한 가상 아이돌 “다테 료코”가 소개된 이후, 디지털 휴먼은 미디어에서 모습을 드러내기 시작했다. 하

지만 이때 당시에는 기술력의 부족으로 인해, <그림 2>와 같이 당시 만들어진 디지털 휴먼은 시청자들에게 불쾌한 경험을 남겼다.

컴퓨터 그래픽의 발전과 모션 캡처 기술의 발명을 통해 지금의 디지털 휴먼은 이전과는 다른 모습을 보이고 있다. 단순히 사람과 유사한 형태를 구현하는 것을 넘어 인간과 컴퓨터의 상호작용에서 새로운 패러다임을 선사하고, 영화나 드라마에서 관객의 몰입감을 극대화하기 위해 특수 효과와 분장만으로는 표현할 수 없는 캐릭터를 연기하며, 사람이 직접 연기하기 어렵거나 위험한 스텐트 장면을 소화하고 있다. 이처럼 디지털 휴먼은 이전엔 사람을 모방하기 위해 연구하는 분야였지만, 점점 사람을 대체하거나 사

<p>ModelingCafe, '이마(Imma)' (가상 인플루언서/모델)</p>	<p>디오비스튜디오, '루이(RUI)' (가상 가수)</p>	<p>EVR스튜디오, 프로젝트 TH(가제) (게임 캐릭터)</p>
		
<p>삼성전자, '네온(Neon)' (고객 안내 서비스 등)</p>	<p>LG전자, '김래아' (제품 홍보 등)</p>	<p>WHO '플로렌스' (금연 상담)</p>
		
<p>IP소프트, '어멜리아(Amelia)' (고객 상담 서비스 등)</p>	<p>Soul Machines, '윌(Will)' (교육 서비스)</p>	<p>머니브레인, '시 아나운서' (방송 서비스)</p>
		

<그림 1> 일상생활에서 볼 수 있는 디지털 휴먼(SPRI Analysis에서 발췌)



<그림 2> 일본의 연예기획사 호리프로에서 만든 다테 쿄코

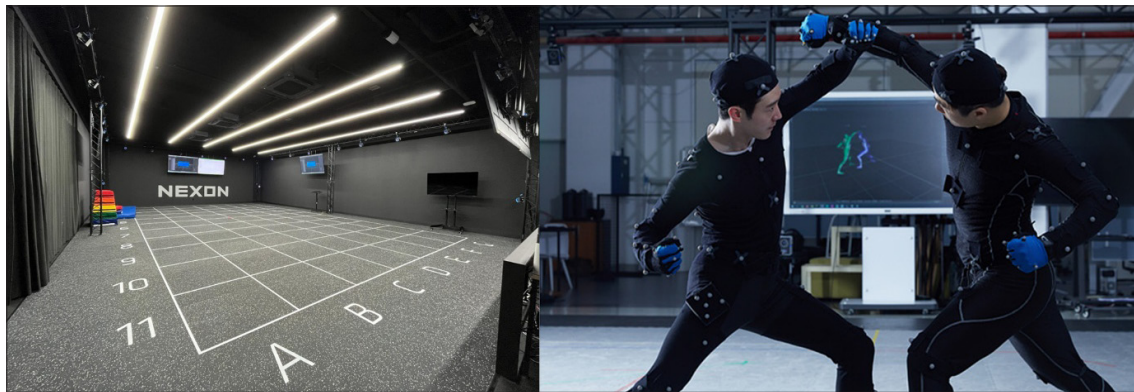
람을 뛰어넘기 위한 기술로 발전하고 있다.

이러한 디지털 휴면을 제작하기 위해서 가장 중요한 기술은 바로 컴퓨터 그래픽과 모션 캡처 기술이다. 컴퓨터 그래픽은 단순한 뼈대 위에 살을 덧붙여 사람 형상을 입체적으로 그리는 기술이다. 뼈대 위에 살을 붙이기 때문에 단순히 뼈대를 움직임으로써 다른 동작을 위해 그림을 다시 그릴 필요가 없다는 장점이 있다. 하지만 컴퓨터 그래픽만으로 디지털 휴면을 만들게 되면 디지털 휴면 하나를 만드는데 소요되는 작업 시간이 많아지고, 실제 사람처럼 움직이게 만들기 어렵다는 문제가 있다.

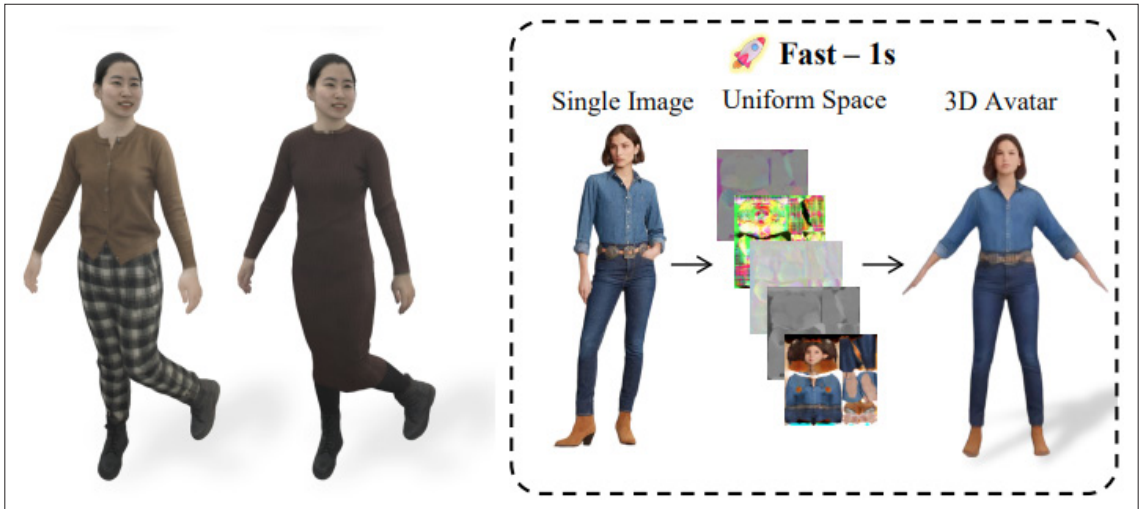
이를 보완하기 위한 기술이 바로 모션 캡처이다. 모션 캡처는 실제 사람의 움직임을 정교하게 추적하여 움직임 정보를 디지털 형태로 수집하는 기술이다. 모션 캡처 기술은 크게 광학식과 관성식으로 분류할 수 있다. 광학식은 수트에 붙은 점들을 <그림 3>에서 볼 수 있는 스튜디오처럼 여러 대의 적외선 카메라가 추적하여 사람의 관절점을 계산하는 방식이다. 촬영자는 단순히 점이 붙어있는 수트를 입고 움직이면 되기 때문에 움직임을 방해하는 장애물이 없고, 여러 대의 적외선 카메라를 통해 추적하기 때문에 정교한 촬영이 가능하다는 장점이 있지만, 적외선 카메라 자체의 비용이 많이 들고 실내에서만 촬영이 가능하다는 단점이 있다.

반대로 관성식은 별도의 카메라 없이 <그림 3>에서 볼 수 있는 장면처럼 수트에 선으로 연결된 IMU센서를 통해 움직임을 추적하는 방식이다. 여기서 IMU센서는 가속도계, 자이로스코프, 지자기 센서가 통합되어 물체의 3차원 움직임을 측정하는 관성 측정 장치이다. 이러한 방식은 많은 장비가 필요하지 않아 상대적으로 저렴하게 촬영이 가능하고, 야외에서도 촬영할 수 있다는 장점이 있지만 상대적으로 정확도가 떨어진다는 단점이 있다.

하지만 최근 디지털 휴면을 위한 기술들은 이러한 단점들을 벗어나려는 시도를 보이고 있다. 생성형 AI와 Diffusion 모델과 같은 인공지능 기술들의 발전으로 기존의 단점들을 해결하려는 것이다. 단순히 사람과 유사하게 생성하는 것을 넘어 <그림 4>의 결과물처럼 입술의 움직임, 피부의 질감과 같은 인체의 다양한 요소들을 최소한의 입력을 통해 복원하려는 것이다. 23년에 발표된 AvatarReX[1]에서는 다중 시점 비디오 입력을 통해 실시간으로 디지털 휴면의 3D모델링을 렌더링하는데 성공하였고, 24년에 발표된 IDOL(Instant Photorealistic 3D Human Creation from a Single Image)[2]에서는 다중 시점을 넘어 단일 시점의 이미지 입력만으로 1초 이내에 전신 3D모델링을 렌더링하는데 성공하였다.



<그림 3> 넥슨의 실내 광학식 모션 캡처 스튜디오(좌)와 펠어비스의 관성식 모션 캡처 장면(우)



<그림 4> AvatarPlex로 만든 디지털 휴먼(좌) 모델과 IDOL로 만든 디지털 휴먼(우)

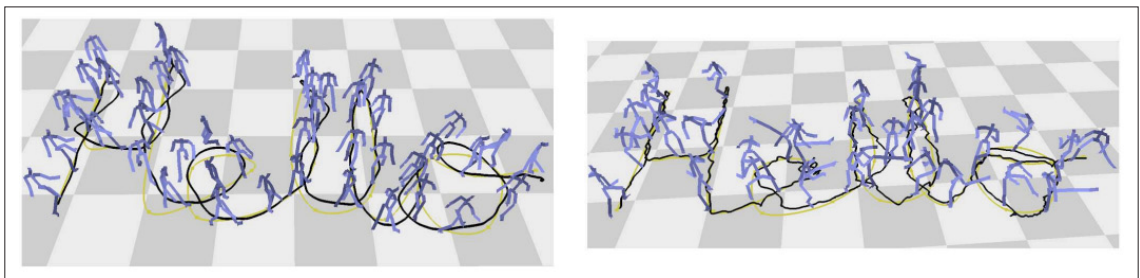
II. 디지털 휴먼 자세 생성

디지털 휴먼의 자세를 생성하는 방식도 변화하고 있다. 초기에는 모션 캡처를 통해 스텐트맨에게 필요한 동작을 모두 지시한 후 수집한 자세를 그대로 사용하거나 비슷한 동작으로부터 역운동학을 계산하여 원하는 자세를 만들어 냈다. 이러한 방식들은 지금도 사용하고 있을 정도로 정확한 자세를 만들어 내는데 적합하지만, 일일이 수동으로 제어하기 때문에 수많은 시간과 비용을 요구한다는 문제점이 있다. 그러나 데이터 확보와 딥러닝이 발전하면서, 이러한 문제점들을 해결하기 위해 수만 개의 자세 데이터를 AI에게 학습시켜 자세를 생성하려는 시도가

이 본격화되었다.

1. 초기 인공지능 신경망의 활용

초기에는 단순히 자세를 생성하는 과제에 집중하였다. 인공지능이 등장하기 이전부터 연구자들은 오랜 시간 누적된 방대한 양의 모션 캡처 데이터를 응용하기 위해 Motion Graph[3]와 같은 연구를 진행해 왔다. Motion Graph에서는 그동안 누적된 수많은 모션 캡처 데이터를 아주 짧은 클립 단위로 잘게 쪼갬 다음, 각 동작이 자연스럽게 이어질 수 있는 지점들을 찾아 수학적인 그래프로 연결하였다. 이러한 방식은 <그림 5>와 같이 기존 데이터에



<그림 5> Motion Graph 알고리즘으로 만든 'HELLO' 글자를 만들며 걷는 동작

서는 존재하지 않은 새로운 연속 동작들을 만들어 내는 획기적인 방법으로 오랫동안 쓰였다.

2010년대 중반 이후 딥러닝 기반 3D 모션 연구가 시작되었고 연속된 동작을 출력하기 위해 시계열 데이터 처리

에 적합한 RNN과 순환신경망을 응용하였다. 그 결과로 ERD(Encoder-Recurrent-Decoder)[4] 모델이 발표되었다. 이 모델은 과거 몇 개의 동작 프레임을 입력 받아, 그 다음에는 어떤 자세가 나올지를 예측하며 <그림 6>의 결



<그림 6> ERD 논문에서 공개한 당시 SOTA 모델과의 결과 비교

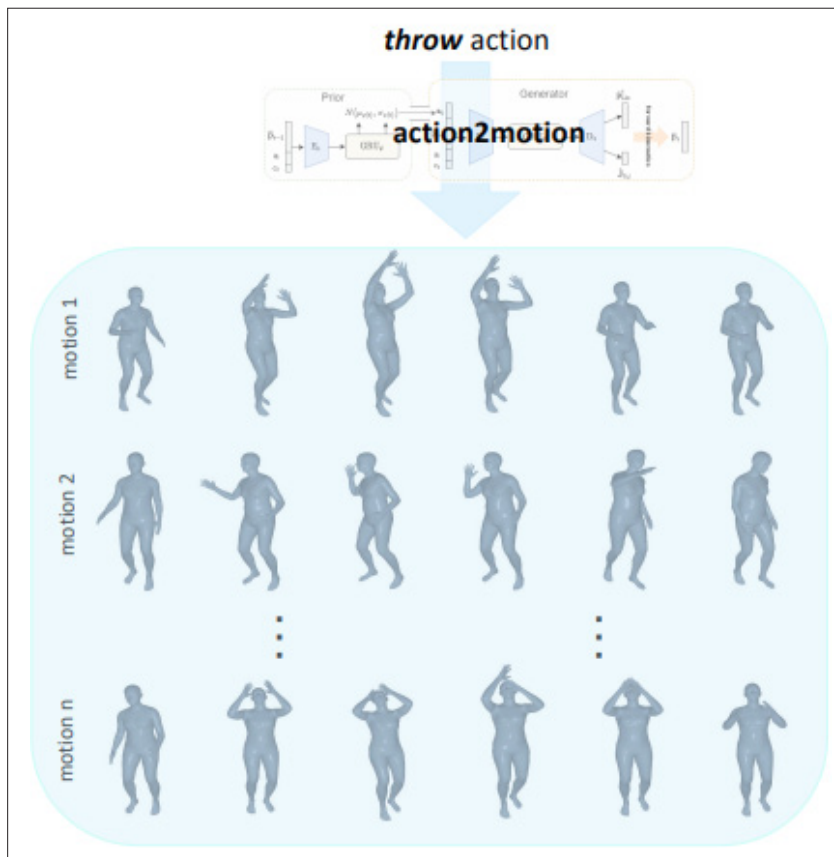
과물처럼 프레임을 하나씩 이어 붙이는 방식으로 동작을 생성했다. 이러한 방식으로 1~2초 정도의 짧은 미래를 예측하여 동작을 생성하는 데는 성공했지만, 순환신경망의 오류 누적 문제를 해결하지 못해 시간이 길어질수록 캐릭터가 굳거나 기괴하게 꺾이는 Drifting 현상이 발생했다.

2. 생성 모델 도입과 제어의 고도화

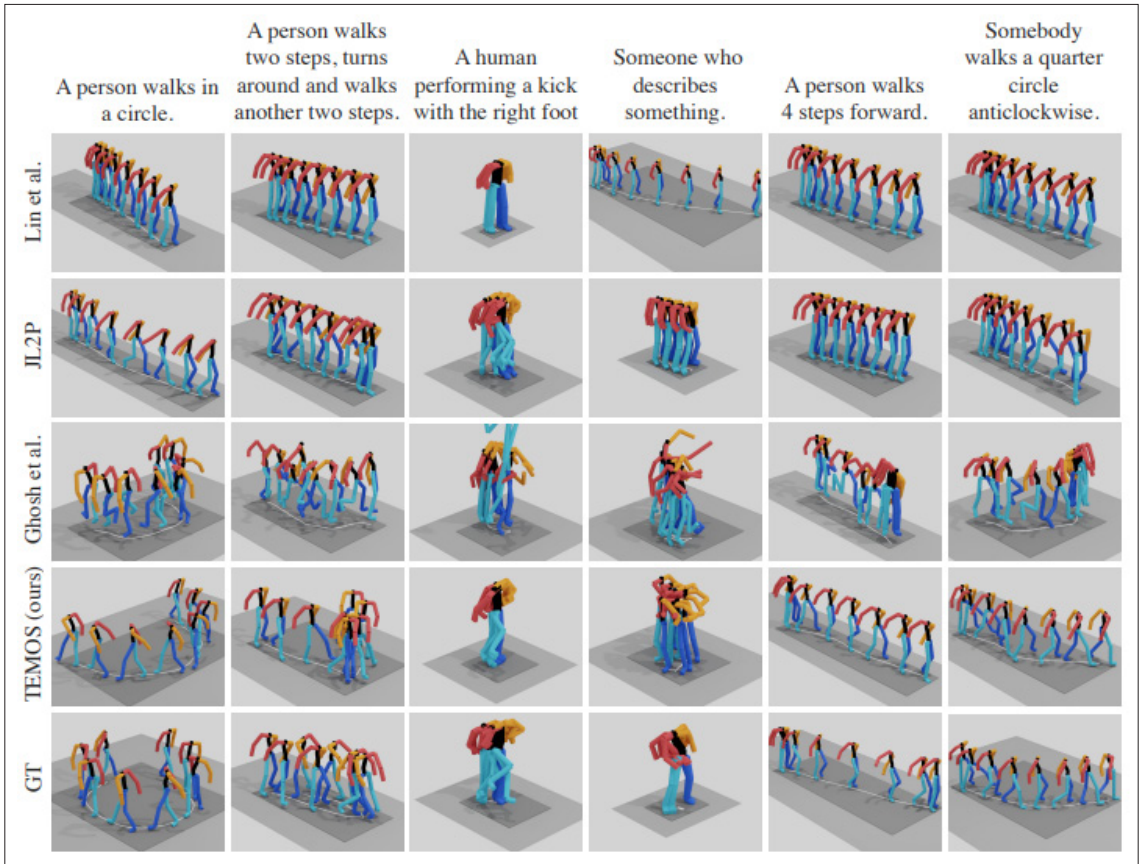
이러한 순환신경망 기반 모델의 문제점을 해결하기 위해, 연구자들은 단순히 다음 프레임을 예측하는 것을 넘어 전체 동작의 시퀀스를 한번에 생성하는 방향으로 연구 방향을 전환하였다. 이를 위해 VAE(변분 오토인코더)나

GAN(적대적 생성 신경망)과 같은 본격적인 생성형 모델들이 연구에 도입되었다.

2020년에 발표된 Action2Motion[5]에서는 VAE 모델을 사용하여 ‘걷기’, ‘뛰기’와 같이 특정 동작들의 라벨을 입력하면 그에 맞는 동작을 생성하도록 설계하였다. 또한 인체 관절의 움직임은 수학적으로 모델링하여 물리적으로 불가능한 자세가 출력되는 현상을 방지하였고, <그림 7>과 같이 매번 다른 동작이 생성되는 확률적 생성의 개념을 도입하였다. 하지만 이러한 기술의 도입에도 불구하고 생성 모션이 길어지면 동작이 무너지는 현상을 막을 수는 없었고, ‘행복하게 팔을 벌리고 걷기’와 같이 세부적인 동작의 묘사와 감정의 표현까지는 묘사할 수 없었다.



<그림 7> Action2Motion 모델의 자세 생성 파이프라인

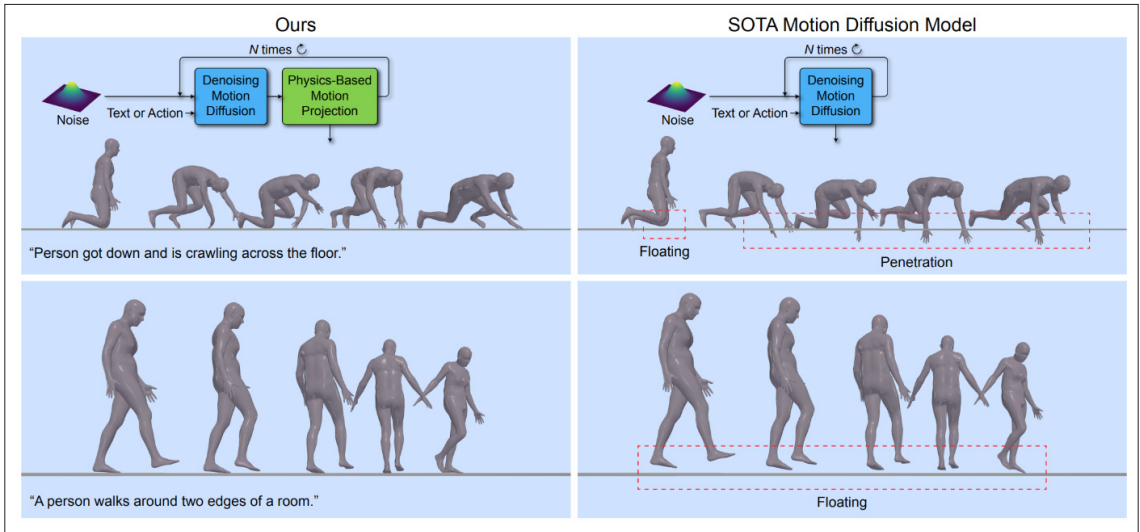


<그림 8> TEMOS 모델의 구체적이거나 추상적인 프롬프트에서의 자세 생성

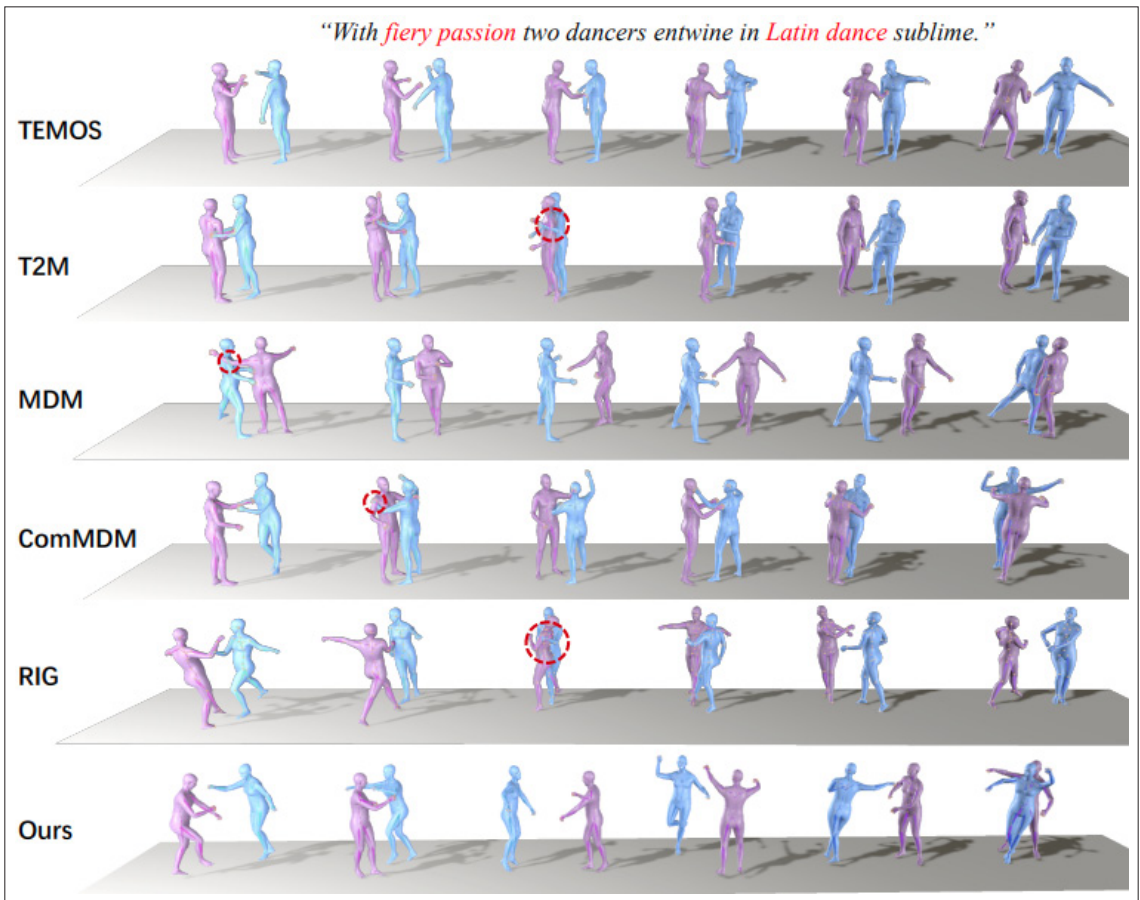
이러한 점들을 개선하기 위해 시계열 데이터와 언어 처리에 탁월한 성능을 보이는 Transformer 구조를 도입하기 시작했고, 그 결과로 TEMOS(Text to MOTions) [6] 모델이 제시되었다. 이 모델은 앞서 언급한 모델과 다르게 VAE 모델과 Transformer 구조를 결합하여 텍스트와 모션을 하나의 잠재 공간에서 통합하도록 설계하였다. 그 결과 언어 모델을 통해 문장의 의미를 추출하고, Transformer를 통해 추출한 3D 동작 데이터의 특징을 하나의 잠재 공간 안에서 정렬하여 <그림 8>의 결과물처럼 다양한 묘사에 대한 적절한 동작을 출력하는데 성공하였다. 또한 Transformer 구조는 데이터의 시간적 흐름을 파악하는 능력이 뛰어나서 순환신경망의 문제였던 Drifting 현상 또한 개선할 수 있었다.

3. 디퓨전 모델의 등장

이러한 동작 생성연구의 발전은 디퓨전 모델의 등장으로 더욱 가속화되고 있다. 최근 연구에서는 더욱 현실적으로 보이는 동작의 생성을 위해, 주변 환경과의 상호작용 또는 다른 디지털 휴먼과의 상호작용을 중점으로 연구가 진행되고 있다. 2023년에 발표된 PhysDiff[7]에서는 <그림 9>의 결과물처럼 그럴듯한 동작의 생성을 넘어 중력과 마찰력과 같은 주변 환경까지 계산하여 동작을 생성하였다. 디퓨전의 생성 과정 안에 일종의 ‘물리 시뮬레이터’를 결합하여 중력에 의해 무게 중심이 무너지지 않는지, 마찰력에 의해 발이 바닥에 고정되는지를 검증하고 이를 계속해서 교정하는 과정을 거친다. 이러한 과정을 통해



<그림 9> PhysDiff 모델에서 해결하고자 한 문제점



<그림 10> 각 모델을 통해 생성한 2인 상호작용 동작 비교

단순히 시각적으로 그럴싸한 동작을 생성하는 것을 넘어, 실제 인간의 무게감과 물리적 제약까지 완벽하게 모방하는 동작을 생성하는 것이다.

이처럼 주변 환경과의 상호작용뿐만 아니라 타인과의 상호작용까지 고려하는 연구 또한 진행 중이다. 24년에 발표된 InterGen[8] 모델은 한 번의 생성 과정에서 두 사람이 서로 상호작용을 하는 동작을 생성하는데 성공하였다. 일반적인 자세 생성 모델을 병렬로 연결하여 자세를 생성하게 되면 <그림 10>의 비교 이미지처럼 각 인물간의 동작과 위치를 파악하지 못해 몸이 겹치거나, 허공에 악수를 하는 등 서로 상호작용을 하지 못하는 동작이 생성된다. 그래서 InterGen 모델에서는 하나의 Transformer가 두 개의 파이프라인을 통해 동작하여 두 사람의 현재 위치와 상호작용을 해야 하는 시점 등을 공유하며 자세를 생성하게 된다.

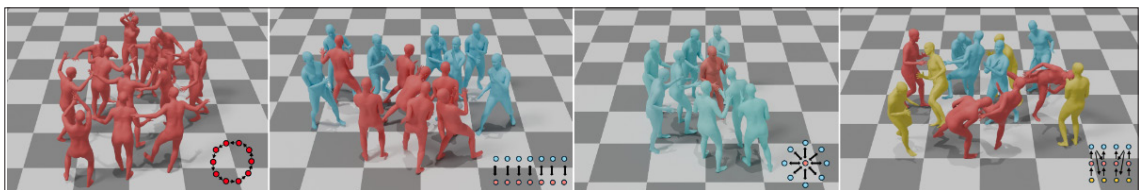
이러한 두 사람 간의 상호작용 생성을 넘어, 최근에는 3인 이상의 다인원이 동시에 상호작용하는 복잡한 환경의 생성으로 연구가 확장되고 있다. 기존에 1인 또는 2인 생성 모델이 학습하는 것처럼 다인원의 자세를 생성하기 위해서는 각 인원에게 맞는 데이터셋을 필요로 하는데, 3인 이상부터는 이러한 데이터셋을 구하기 매우 힘들다는 문제점이 있다. 이 문제를 해결하기 위한 연구가 25년 발표된 Multi-Person Interaction Generation from Two-Person Motion Priors[9] 논문이다. 여기서는 다인원의 상호작용 관계를 작게 분할하여 ‘2인 상호작용 관계들의 집합’으로써 간주하였다. 이러한 변환을 중심으로 다인원 상호작용을 두 사람 단위의 관계 그래프로써 설계한다. 이러

한 그래프 구조를 바탕으로 디퓨전 모델이 동작을 생성하는 과정에서 각 인물 간의 위치와 움직임을 조정하게 되고, <그림 11>의 결과물처럼 인물들의 몸이 겹치거나 충돌하는 물리적 오류 없이 3인 이상의 자연스러운 그룹 상호작용을 생성하게 된다. 이러한 방법으로 막대한 시간이 필요한 초기 학습과정과 막대한 비용을 요구하는 데이터셋 구축 없이 여러 사람의 자연스러운 동작을 고품질로 생성할 수 있게 되었다.

III. 디지털 휴먼 리타게팅 기술

자세 생성 단계에서 고도화된 자세 생성 모델을 통해 아무리 자연스럽고 정교한 동작을 만들어 내더라도, 이를 실제 산업에 적용하기 위해서는 모션 리타게팅 기술이 필요하다. 리타게팅이란 원본이 되는 캐릭터의 관절점 움직임이나 관절의 길이 같은 데이터를 체형과 구조가 다른 타겟 캐릭터에 맞게 변환하고, 이를 이식하는 기술이다.

소셜 미디어, 메타버스, 게임, 영화 및 드라마 같은 현대의 3D 산업에서는 평균적인 체형의 사람만 나타나는 것이 아니다. 다양한 체형과 신체적 특징을 가지는 사람, 로봇, 또는 가상의 종족 및 괴물 등 수많은 형태의 디지털 캐릭터가 존재하고, 같은 캐릭터라도 내부 관절점은 다른 형태로 구성될 수 있다. 매번 각 캐릭터의 체형과 관절 정보에 맞춰 모션을 새로 캡처하거나 생성하는 것은 막대한 시간과 비용을 초래하므로, 하나의 고품질 모션 데이터를 다양한 골격에 이식할 수 있는 리타게팅 기술은 디지털 휴



<그림 11> 다인원의 상호작용 관계 그래프와 동시에 생성한 자세

면 제작을 위한 파이프라인에서 중요한 역할을 담당한다.

1. 과거의 리타게팅 기술

초기의 리타게팅 연구는 주로 수학적 연산을 통한 최적화와 운동학을 기반으로 이루어졌다. 1998년 발표된 Gleicher의 연구[10]에서는 <그림 12>에서 순차적으로 묘사하듯이 캐릭터의 각 관절의 비율이 달라지더라도 ‘발은 바닥에 닿아 있어야 한다.’, ‘손목의 관절은 특정 각도를 유지해야 한다.’와 같은 식의 시공간적인 제약 조건을 설정하였고, 이를 만족하는 관절의 회전각을 역운동학을 통해 계산하는 방식을 제안했다. 비록 단순하고 간단한 조건을 추가하는 연구였지만, 비슷한 관절을 가지는 캐릭터에게 자세를 이식하는 데에는 성공한 시도였다. 하지만 이 방식은 역운동학을 연산하는 과정이 매우 복잡하고, 타겟과 원본의 체형이 크게 다를 경우 관절이 기괴하게 꺾이거나 원래 동작의 의미를 상실할 수 있다는 한계점이 존재했다.

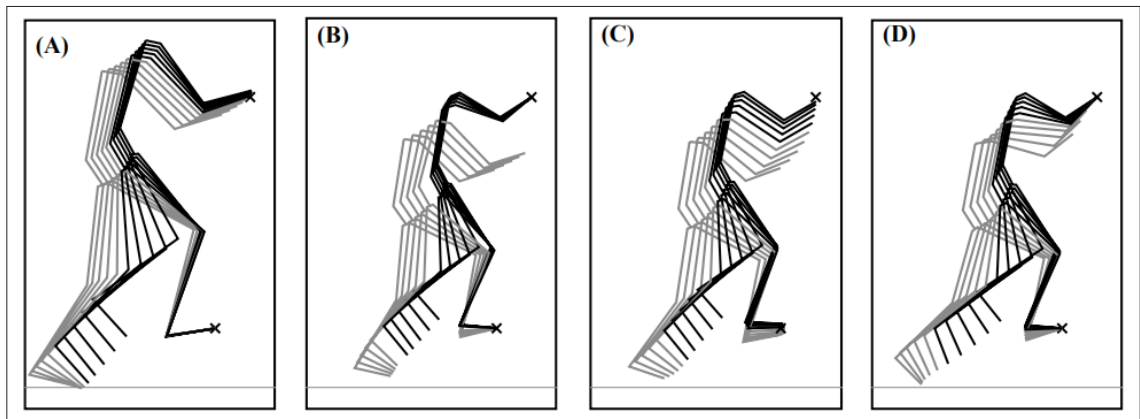
이런 복잡한 연산의 속도를 개선하기 위해 2005년 발표된 A Physically-Based Filter for Geometrically-Generated Motion[11] 논문에서는 로봇의 자세 제어에 쓰이던 동역학의 ZMP(Zero Moment Point)를 통해 관

절의 한계와 균형 유지를 필터 형태로 만들어 실시간으로 계산하는 방식을 제안하였다. 또한 2008년에는 게임 산업의 발전과 함께 사람과는 골격 형태가 다른 캐릭터에게 자세를 이식할 필요성이 늘어났고, 이를 해결하기 위해 Real-time Motion Retargeting to Highly Varied User-Created Morphologies[12]와 같은 연구가 진행되었다. 이 연구는 사전에 수많은 제약 조건들을 수학적으로 프로그래밍해 두고, 관절이 다른 캐릭터의 뼈대 구조에 맞춰 실시간으로 재구성하는 방법을 제안하였다.

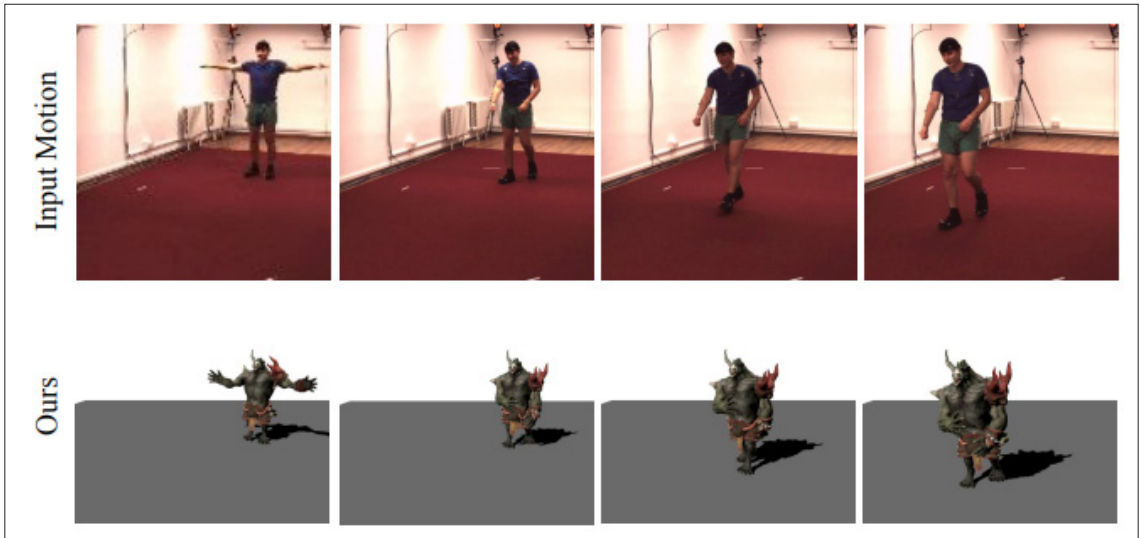
하지만 이러한 방법들에도 불구하고, 결국은 사람이 직접 복잡한 물리법칙을 세팅하고, 무수히 많은 수학적 제약 조건을 프로그래밍해야 한다는 치명적인 문제가 존재했다. 이러한 엔지니어링 한계는 2010년대 딥러닝이 본격적으로 도입되며 마침내 새로운 돌파구를 찾게 된다.

2. 딥러닝을 통한 리타게팅

2018년 발표된 NKN(Neural Kinematic Network)[13]는 인공지능망 내부에 인체의 관절점의 위치를 계산하는 순운동학 계층을 직접 설계하여 결합한 모델이다. 단순히 인공지능망을 사용한 모델에서는 관절의 3D 좌표 자체를 예측하려고 했기 때문에 학습 과정에서 관절의 길이가 늘



<그림 12> 제약 조건을 위한 고정점과 역운동학을 통한 리타게팅 과정



<그림 13> 사람의 움직임을 NKN 모델을 통해 가상의 캐릭터에 이식한 결과물

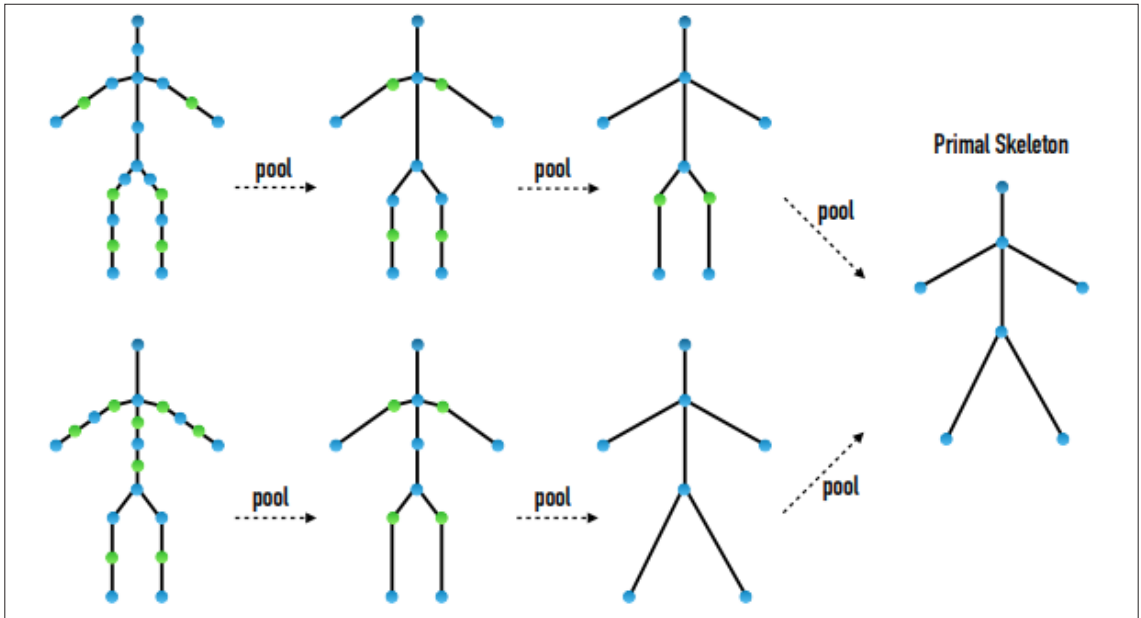
어나거나 줄어드는 물리적인 오류가 빈번하게 발생했다. 하지만 NKN에서는 인공신경망 안에 순운동학 계층을 내장함으로써 인공지능은 단순히 관절의 각도만 예측하고, 순운동학 계층에서 타겟의 관절 길이를 고정된 채로 최종 위치를 계산하도록 하였다. 이를 통해 딥러닝 모델이 생성한 위치가 물리적으로 왜곡되는 현상을 차단하였다. 그 결과 <그림 13>과 같이 사람의 움직임을 가상의 캐릭터에 그대로 이식하는 것이 가능해졌다.

또한 NKN에서는 데이터의 문제까지 해결하였다. 일반적으로 리타게팅 모델을 학습시키기 위해서는 체형이나 관절이 다른 두 캐릭터가 완전히 똑같은 동작을 수행한 1:1로 매칭된 데이터가 필수적이다. 하지만 이를 구하는 것은 거의 불가능에 가깝다는 문제가 있었다. 그러나 NKN에서는 순환 일관성이라는 개념을 도입하여 이러한 문제를 해결하고자 했다. 이는 원본 캐릭터의 동작을 타겟에서 이식했다가 다시 원본 캐릭터에게 이식하는 방식이다. 이 방식을 통해 모델은 입력된 동작과 순환한 동작의 오차를 줄이도록 학습하기 때문에 현실적으로 구할 수 없는 정답 데이터 없이도 자연스럽게 모션을 전이하도록 학습하는 것이다.

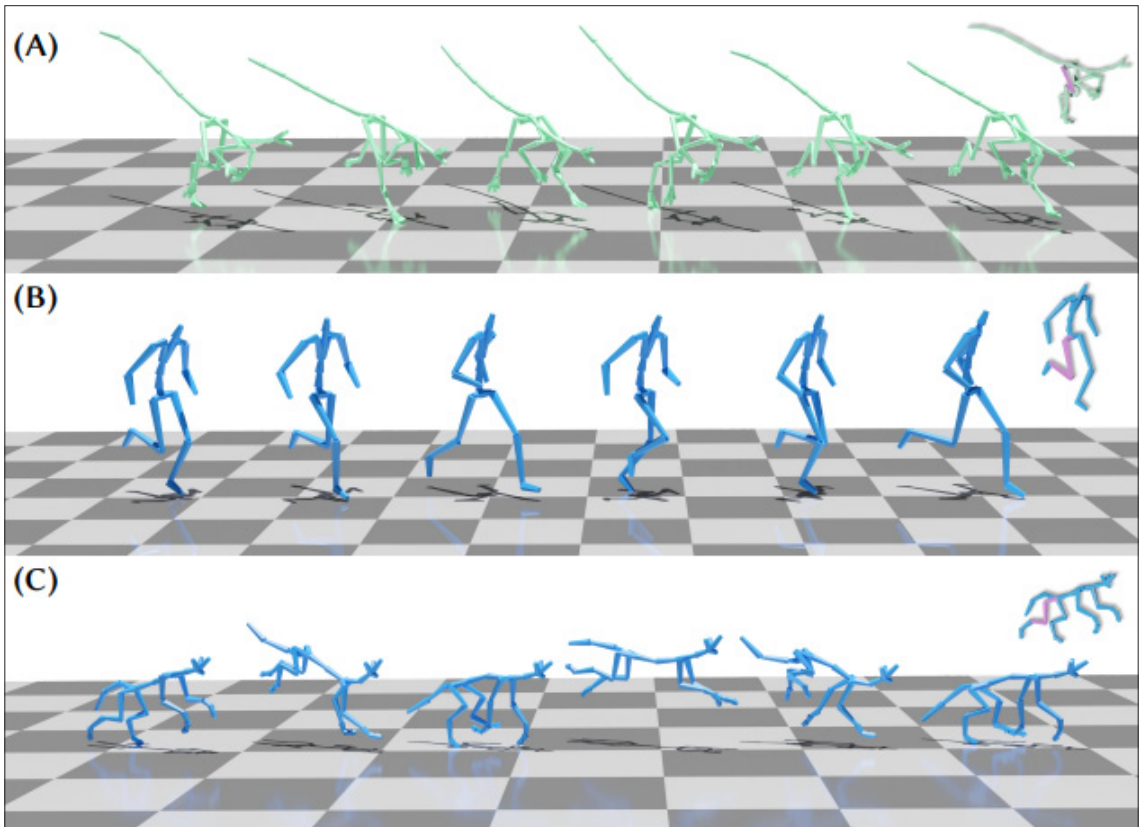
하지만 이러한 획기적인 방식에도 불구하고 초기 딥러닝 모델들은 치명적인 문제점이 있었다. 관절의 길이가 다른 경우에는 원활하게 자세를 이식할 수 있었지만, 전체 관절의 개수나 연결구조가 다른 경우에는 이식이 불가능하다는 것이다. 이러한 이종 골격 간의 리타게팅 문제를 해결하기 위해 2020년에 SAN(Skeleton-Aware Networks)[14] 모델이 연구되었다.

이 논문에서는 구조가 각기 다른 뼈대들의 움직임을 통일시키기 위해 <그림 14>의 과정을 통해 원시 뼈대라는 공통된 잠재 공간을 도입하였다. 이 원시 뼈대를 통해 모델은 동작의 순수한 의미와 맥락만을 추출하는 것이다. 이렇게 추출된 동작의 의미를 타겟 캐릭터에 맞춰 다시 풀어내는 방식을 취함으로써, 체형이 다른 캐릭터뿐만 아니라 관절 개수와 구조가 완전히 다른 타겟에게도 자세를 이식하는데 성공하였다.

최근에는 리타게팅 연구 역시 Diffusion 모델과 결합하며 다른 모습을 보이고 있다. 25년에 발표된 Motion2Motion[15]에서는 사전학습된 Diffusion 모델을 통해 리타게팅을 진행하였다. 이 연구에서는 원본 캐릭터와 타겟 캐릭터의 손끝, 발끝과 같이 아주 적은 수의 핵심 관절들



<그림 14> 다른 관절점의 캐릭터를 공통된 원시 뼈대로 만드는 과정



<그림 15> 관절 대응을 통해 랩터(A)의 동작을 사람(B)과 원숭이(C)에 이식한 결과

을 대응시키면, 디퓨전 모델이 노이즈를 제거하는 단계에서 나머지 지정되지 않은 관절들의 위치를 물리적으로 적합한 위치에 생성하며 채우는 것이다. 이러한 방식을 통해 <그림 15>와 같이 별도의 추가 학습과정 없이 공룡의 움직임을 사람이나 원숭이와 같은 완전히 다른 종족에게도 이식하는데 성공하였다.

IV. 결론

본 논고에서는 디지털 휴먼을 제작하기 위해 필요한 기술 중 자세 생성과 자세 리타게팅 기술들의 연구동향 및 최신 발전 과정을 상세히 살펴보았다. 단순히 자세를 생성하는 것을 넘어 주변 환경이나 다른 디지털 휴먼과의 상호작용을 이해하며 생성된 자세를, 이중 골격 간의 제약을 극복하고 별도의 추가학습 없이 동작을 이식하는 과정을 통해 고품질의 디지털 휴먼 제작 파이프라인이 완성된다. 이는 과거 수학적 연산과 수작업에 의존하던 물리적-구조적 한계를 극복하고, 가상세계에 더욱 현실성을 불어넣을 수 있는 핵심 기반이 된다.

앞으로 이러한 기술들이 더욱 발전되어 현실을 이해하

고 더 많은 객체들과 상호작용할 수 있을 정도로 정교해지면, 디지털 휴먼의 활용범위는 지금보다 훨씬 더 넓어질 것으로 보인다. 게임과 메타버스 생태계를 넘어, 가상에서 생성한 디지털 휴먼을 현실의 휴머노이드에 이식하는 로봇 공학 분야까지 같이 성장할 수 있고, 이미 멸종된 생명체들을 가상현실에서 복원하거나 인간이 직접 행동하기 어려운 동작들을 디지털 휴먼이 대체함으로써 영화나 드라마에서 더욱 현실적인 묘사가 가능할 것이다. 나아가 모션 캡처 스튜디오나 전문적인 애니메이션 편집 기술 없이도, 개인이 간단한 텍스트 지시만으로 고품질의 3D 콘텐츠와 디지털 휴먼을 직접 제작하여 온라인 콘텐츠를 제작하는 단계까지 도달할 수 있다.

하지만, 이러한 기술들이 발전함에 따라 무용수나 댄서들의 고유한 움직임을 무단으로 모방하거나 학습하는 동적 저작권 침해 문제나, 미세한 움직임까지 완벽히 모방하는 3D 딥페이크 악용과 같은 사회적 문제를 야기할 가능성이 있다. 이러한 사회적 문제들을 간과하지 않고 기술적인 보안과 윤리적인 가이드라인을 함께 마련한다면, 디지털 휴먼 기술은 개인이 금전적/기술적 문제를 떠안을 걱정 없이 새로운 형태의 미디어를 창작/소비하는 새로운 세상을 만들어 낼 것으로 기대된다.

참 고 문 헌

- [1] J. Zheng, et al., "AvatarReX: Real-time Expressive Full-body Avatars," *ACM Transactions on Graphics (TOG)* 42 (2023), pp. 1-15
- [2] Yiyu Zhuang, et al., "IDOL: Instant Photorealistic 3D Human Creation from a Single Image", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 26308-26319
- [3] L. Kovar, M. Gleicher, and F. Pighin, "Motion Graphs," *ACM Transactions on Graphics (TOG)* 21 (2002), pp. 473-482
- [4] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent Network Models for Human Dynamics," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 4346-4354
- [5] C. Guo, et al., "Action2Motion: Conditioned Generation of 3D Human Motions," *Proceedings of the 28th ACM International Conference on Multimedia (ACM MM)* (2020), pp. 2021-2029
- [6] M. Petrovich, M. J. Black, and G. Varol, "TEMOS: Generating diverse human motions from textual descriptions," *European Conference on Computer Vision (ECCV)* (2022), pp. 480-497

참 고 문 헌

- [7] Y. Yuan, et al., “PhysDiff: Physics-Guided Human Motion Diffusion Model,” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)*, pp. 16010-16021
- [8] H. Liang, et al., “InterGen: Diffusion-based Multi-human Motion Generation under Complex Interactions,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)*, pp. 1494-1504
- [9] W. Xu, S. Fan, P. Henderson, and E. S. L. Ho, “Multi-Person Interaction Generation from Two-Person Motion Priors,” *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH Conference Papers) (2025)*, pp. 113:1-113:11
- [10] M. Gleicher, “Retargetting Motion to New Characters,” *Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH) (1998)*, pp. 33-42
- [11] S. Tak and H. Ko, “A Physically-Based Filter for Geometrically-Generated Motion,” *ACM Transactions on Graphics (TOG) 24 (2005)*, pp. 98-105
- [12] C. Hecker, et al., “Real-time Motion Retargeting to Highly Varied User-Created Morphologies,” *ACM Transactions on Graphics (TOG) 27 (2008)*, pp. 1-11
- [13] R. Villegas, et al., “Neural Kinematic Networks for Unsupervised Motion Retargetting,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)*, pp. 8639-8648
- [14] K. F. Aberman, et al., “Skeleton-Aware Networks for Deep Motion Retargetting,” *ACM Transactions on Graphics (TOG) 39 (2020)*, pp. 62:1-62:14
- [15] L.-H. Chen, Y. Zhang, Z. Yin, Z. Dou, X. Chen, J. Wang, T. Komura, and L. Zhang, “Motion2Motion: Cross-topology Motion Transfer with Sparse Correspondence,” *Proceedings of the SIGGRAPH Asia 2025 Conference Papers (2025)*, pp. 1-11

저 자 소개



조경민

- 2019년 3월 ~ 2025년 2월 : 경상국립대학교 메카트로닉스공학부 메카트로닉스공학 학사
- 2025년 3월 ~ 현재 : 서울과학기술대학교 국방융합과학대학원 국방ICT융합공학과 석사과정
- ORCID : <https://orcid.org/0009-0006-5977-7190>
- 주관심분야 : 피지컬 AI, 딥러닝, 데이터 처리



강현석

- 2015년 3월 ~ 2021년 8월 : 강원대학교 삼천캠퍼스 전자공학과 학사
- 2021년 9월 ~ 2023년 8월 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 석사
- 2023년 9월 ~ 현재 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 박사과정
- ORCID : <https://orcid.org/0000-0003-0783-3841>
- 주관심분야 : 3D 컴퓨터 비전, 딥러닝

저 자 소 개



이 정 우

- 2020년 3월 ~ 현재 : 서울과학기술대학교 전자미디어공학과 학사
- ORCID : <https://orcid.org/0009-0007-0582-6185>
- 주관심분야 : 딥러닝, 컴퓨터 비전



박 구 만

- 1984년 : 한국항공대학교 전자공학과 공학사
- 1986년 : 연세대학교 대학원 전자공학과 석사
- 1991년 : 연세대학교 대학원 전자공학과 박사
- 1984년 ~ 1996년 : 삼성전자 신호처리연구소 선임연구원
- 1991년 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- 2006년 ~ 2007년 : Georgia Institute of Technology, Dept. of ECE, Visiting Scholar
- 2016년 ~ 2017년 : 서울과학기술대학교 나노IT디자인융합대학원 원장
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터 비전, 실감미디어