

메타버스 서비스를 위한 휴먼 모델링 기술 동향

□ 박민규, 강주미, 윤주홍 / 한국전자기술연구원

요약
영상, 비디오 정보로부터 3D 휴먼 모델을 생성하고 실제 사람 같이 움직이게 변형하는 기술은 다가오는 메타버스 시대의 핵심 기술이며, 이와 같은 기술이 갖는 잠재적인 파급력 때문에 많은 글로벌 IT 기업에서 앞다투어 관련 기술에 대한 투자를 진행하고 있다. 본 기고문에서는 지난 10년간 휴먼 모델링 기술이 어떻게 발전되어 왔는지 살펴보고 현재 기술의 특징 및 한계점과 함께 앞으로의 기술 전망에 대해 살펴본다.

I. 서론

휴먼 모델링 기술은 지난 30여 년간 다양한 연구기관, 학교, 기업에서 폭넓게 연구되어온 주제 중 하나이며, 기술이 성숙됨에 따라 광고, 영화, 뮤직 비디오[1,2] 등 다양한 산업에서 새로운 형태의 콘텐츠 제작을 위해

활용하기 시작하였다. 하지만 정교한 3D 휴먼 모델을 복원하기 위해서는 많은 수의 카메라와 방대한 컴퓨팅 자원이 요구되며, 필요에 따라서는 사실성을 높이기 위해 짧게는 수 일에서 길게는 수 개월간 숙련된 전문가의 후작업을 필요로 한다. 또한, 상용화 수준의 복원은 일반적으로 대규모 렌더팜을 활용한다. 따라서 앞서 언급한 시간적, 공간적, 경제적 한계점으로 인해 일반 사용자 보다는 배우, 가수 등 유명인이 출연하는 고가의 콘텐츠 제작을 위해 주로 사용이 되고 있다.

이와 같이 유명인의 3D 모델링을 기반으로 하는 서비스는 초창기 메타버스의 형태라고 할 수 있으며, 최근 개최되고 있는 가상 콘서트[4,5] 등이 여기에 해당한다. 다음 단계의 메타버스 서비스를 위해서는 특정 인물이나 유명인에 국한되지 않고 누구나 쉽고 빠르게 자신의 아바타를 메타버스 공간 안에 생성하고 제어할 수 있는

※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00103, 가상공간구성을 위한 5G 기반 3D 공간 스캔 디바이스 기술 개발)

기술 개발이 선행되어야 하며 이를 위해서는 고가의 스튜디오 환경이 아닌 단일 카메라나 저가의 웹캠 카메라 기반의 3D 휴먼 모델링 기술 개발이 선행되어야 한다. 궁극적으로는 일반 사용자에게 대한 디지털화뿐만 아니라 반려 동물이나 사물, 공간에 대한 복원까지도 이루어져야 하고, 스마트 폰이나 웹캠 등을 통해 일반적인 가정 환경에서도 손쉽게 사람과 사물의 디지털화가 이루어질 수 있어야 진정한 메타버스 시대를 맞이할 수 있을 것이라 생각이 된다.

다행히 딥 러닝 기술의 발전으로 인해 3D 모델링 기술은 매우 빠르게 발전되고 있으므로 머지 않은 시기에 메타버스 공간에서 자신의 아바타를 이용하여 가상의 공간을 경험하는 시대가 올 것으로 예상된다. 본 기고문에서는 전통적 스튜디오 기반의 복원 기술, 단일 영상/웹캠 카메라 기반의 복원 기술, 템플릿 모델 생성/변형 기술, 뉴럴 렌더링 기술 등 메타버스 서비스를 위한 다양한 휴먼 모델링 기술에 대해 살펴보고 현재 당면한 문제점과 앞으로의 전망에 대해 살펴보고자 한다.

II. 휴먼 모델링 기술 개요

휴먼 복원 기술은 크게 스튜디오 환경에서 다중 시

점 카메라를 통해 획득한 영상을 이용해 복원하는 방식, 다중 깊이 센서를 이용하는 방식, 그리고 단일 카메라나 깊이 센서를 이용하는 방식으로 구분될 수 있다. 그리고 수행 속도에 따라 실시간, 비실시간 기술로 분류하거나, 템플릿 모델 사용 여부에 따라 템플릿 기반 방법, 비모수적(non-parametric) 접근 방법 등으로 구분할 수 있다. 최근 기술이 발전됨에 따라 서로 다른 기술의 장점을 융합하는 방법이 활발히 연구되고 있어 접근 방법을 명확하게 구분짓기는 어려운 부분이 있으므로 본 기고문에서는 편의상 4가지 기술 군으로 분류하여 소개하고자 한다.

1. 스튜디오 기반의 휴먼 복원 기술

딥 러닝 이전의 복원 기술은 폐색 영역(occlusion)에 대한 복원이나 텍스처가 부족한 영역에 대한 복원이 가장 주요한 문제였기 때문에 영상 센서의 개수를 늘려 폐색 영역을 최소화하고 영상 내 대응점 탐색 성능을 높이는 방향으로 기술이 발전되었다. 초창기라 할 수 있는 1990년대와 2000년대 초에는 크로마키 환경에서 전배경 분리를 통해 얻어진 분할 지도를 이용한 복원 방식이 주를 이루었다. 그 후 카메라와 컴퓨팅 자원의 발전과 더불어 컴퓨터 비전 알고리즘도 함께 발전하여 복원 성



<그림 1> 가수 청하(좌: LGU+[1])와 위클리(우:SKT[2]) 복원 장면 및 결과

<표 1> 스튜디오 기반 복원 솔루션 비교

	Microsoft (Mixed Reality Capture Studio)	8i Studio	Intel Studio	Canon Video Studio
복원 대상	1-4인	1-4인	50명 이상 (300평 규모)	최대 10인 (8mx8m)
카메라 대수	106	24-60	100+	100+
해상도	2K+	2-4K	8K	4K
특징	IR 구조광 기반 깊이 추정 및 컬러 영상 융합	컬러 영상 기반	뮤지컬 및 영화 제작 목적	스포츠, 공연 복원 목적

능이 크게 향상되었다. 2010년대에 이르러 관련 기술이 솔루션 형태로 판매되기 시작하였으며 대표적으로 미국 Microsoft 사의 혼합 현실 캡처 스튜디오[8], 뉴질랜드의 8i 스튜디오[27], 미국 Intel 사의 Intel Studio[28], 캐논 사의 비디오 스튜디오[29] 등이 있으며 각 솔루션의 특징을 요약하면 다음과 같다.

Microsoft사의 볼류메트릭 스튜디오는 106대의 산업용 카메라로 구성[8]되어 있으며, 획득된 영상은 깊이 추정, 전배경 분할, 깊이 융합, 텍스처 생성 과정 등을 거쳐

최종적으로 3D 모델링이 이루어진다. 다른 기업의 방식과 다르게 적외선 기반의 구조광(structured light)을 사용하여 깊이 정보를 추정한다. 이와 함께 컬러 영상 기반의 깊이 추정 기술, 실루엣 정보를 이용한 깊이 추정 결과를 융합하여 고품질의 3D 모델을 복원하는 접근 방법을 갖는다. 추가적으로 얼굴 부분에 대한 품질 향상을 위한 별도의 텍스처 맵 생성 기술을 제안하였다. 8i 스튜디오의 특징은 컬러 영상만을 이용하여 모델을 복원한다는 것이며, 세밀한 복원을 위해 공간에 대한 사



<그림 2> Microsoft사의 Mixed Reality Capture Studio와 카메라 구성

전 지식을 바탕으로 다양한 초점 거리를 갖는 카메라를 배치한다는 특징이 있다. 예를 들어 얼굴 영역을 촬영하는 카메라는 초점 거리가 큰 카메라를 사용하여 (예: 25mm) 얼굴 영역을 크게 촬영하여 복원하고, 외형 변화가 몸통의 경우 초점 거리가 작은 카메라를 사용하여 보다 넓은 영역을 복원한 뒤, 최종적으로 다양한 품질의 깊이 영상을 융합하는 접근 방법을 갖는다. Intel Studio는 대규모 공간에 대한 복원을 제공한다는 특징이 있는데 8K급 카메라 100대 이상을 사용하여 복원을 수행한다는 특징이 있다. 이를 통해 획득되는 데이터의 양이 수 분 내에 테라바이트에 이르기 때문에 많은 양의 데이터를 효과적으로 처리하고 복원을 수행하는 것이 핵심 기술이라 할 수 있다. 최근에는 Canon사에서 스포츠나 공연 복원을 위한 Canon Video Studio를 선보였으며 해당 기술은 빠른 속도로 움직이는 다중 객체에 대한 복원에 초점을 맞추고 있다.

Microsoft사에서는 동일 시기(2015)에 볼류메트릭 스튜디오와 함께 Fusion4D 기술[17], 그리고 Holoportation 기술을 발표하였는데 끝으로 해당 기술에 대해 간략히 소개를 하겠다. Fusion4D 기술은 일반적인 가정에 약 24대의 카메라를 설치하고, 실시간 깊이 추정 및 융합을 통해 인간을 실시간으로 복원하고 원격지에 있는 사람의 홀로렌즈에 증강해주는 기술로 아직 기술이 상용화되지는 않았지만 앞으로 이와 같은 기술이 상용화될 것으로 예상된다.

2. 딥 러닝 기반의 휴먼 복원 기술

휴먼 모델링 기술은 딥 러닝 기술의 발전과 함께 빠르게 발전되고 있으며 본 절에서는 대표적인 딥 러닝 기반 기술에 대해 살펴보도록 한다. 딥 러닝 기술의 경우 주로 단일 시점에서의 영상 정보나 깊이 영상을 이용하여 전체 모델을 복원 하는 방향으로 기술이 발전되고 있으

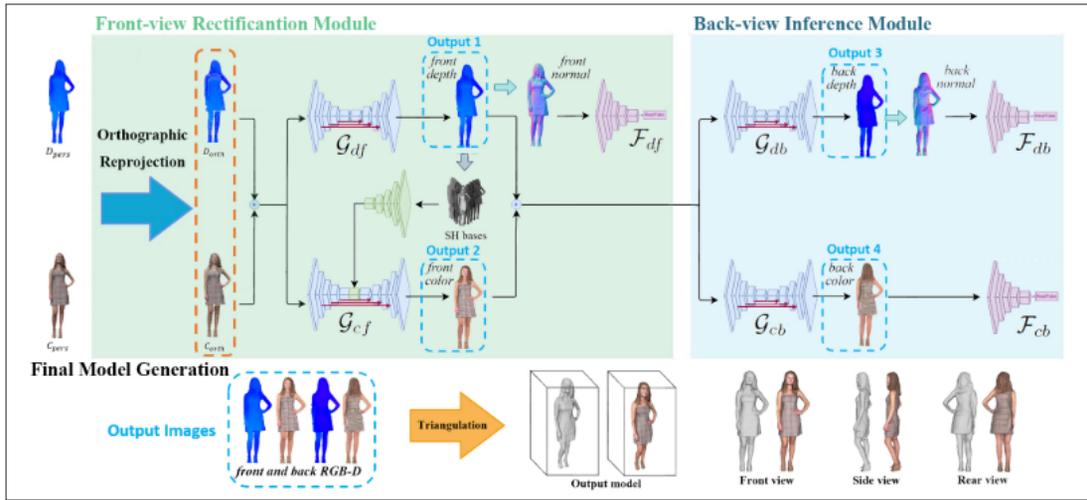
며, 최근에는 비디오 정보를 이용하는 방법에도 연구도 활발히 진행되고 있다.

1) 깊이 센서 기반의 휴먼 모델링 기술

Intel의 RealSense[30], Microsoft의 Kinect[31], Stereo Labs의 ZED[32]와 같이 저가의 범용 태스 카메라가 보급화됨에 따라 단일/다중 깊이 센서 정보를 이용하는 휴먼 및 객체의 3D 모델링 연구가 활발하게 진행되고 있다. 깊이 정보를 이용하여 사람과 같이 움직이는 물체를 복원하는 방법은 크게 단일 깊이 영상을 이용하여 복원하는 방법, 연속된 깊이 영상을 정합하여 점진적으로 휴먼 모델을 복원하는 방법, 그리고 다중 깊이 센서 정보를 사용하여 휴먼 모델을 복원하는 방법으로 나뉘어 질 수 있다.

그 중 연속된 깊이 영상을 정합하면서 점진적으로 복원을 수행하는 방법이 가장 먼저 개발되었으며, 특히 2015년에 발표된 DynamicFusion[18]은 정적 객체가 아닌 동적 객체에 대한 복원이 실시간으로 처리가 가능함을 보여주었다. 이 후 DeepHuman[23], DoubleFusion[25], BodyFusion[26] 등으로 Dynamic Fusion 기술이 확장되었으며 초반에는 주로 Dynamic Fusion이 갖는 한계점을 극복하는 방향으로 연구가 진행되었다. 예를 들어 BodyFusion[26] 연구에서는 빠르게 움직이는 객체에 대한 복원을 가능하게 하기 위해서 인체 관절에 대한 예측 정보를 추가적으로 이용하는 접근 방법을 제안하였고, HybridFusion[24]에서는 관성 센서 정보를 동시에 이용하는 방법을, DoubleFusion[25]에서는 SMPL과 같은 템플릿 모델을 같이 이용하여 움직임 변화에 강건한 복원이 이루어 질 수 있도록 딥 러닝 네트워크 구조를 제안하였다. 지면의 한계로 관련 기술에 대한 구체적인 소개는 방송미디어 학회지 26권 3호[11]를 참조하기 바란다.

한편으로는 연속된 영상간 정합을 통한 점진적 복원



<그림 3> NormalGAN[12]: RGB-D 영상을 입력으로 받아 음영이 제거된 영상과 후면에 대한 깊이 영상과 컬러 영상을 예측하는 기술

이 복잡도가 높고 최초 복원되는 원형 모델(canonical model)에 대한 안정적 업데이트가 어렵다는 한계점 때문에 단일 깊이 영상만을 이용해서 전신에 대한 복원을 수행하고자 하는 시도가 자주 이루어지고 있다. NormalGAN[12]의 경우 정면에서 촬영된 깊이 영상으로부터 후면 깊이 영상을 예측하는 접근 방법을 제안하였으며 보다 좋은 결과를 생성하기 위해 입력 영상의 투영 성분을 제거하는 방법과 음영을 제거하는 방법, 법선 지도에 대한 식별자(discriminator)를 이용하는 방법을 제안하였다.

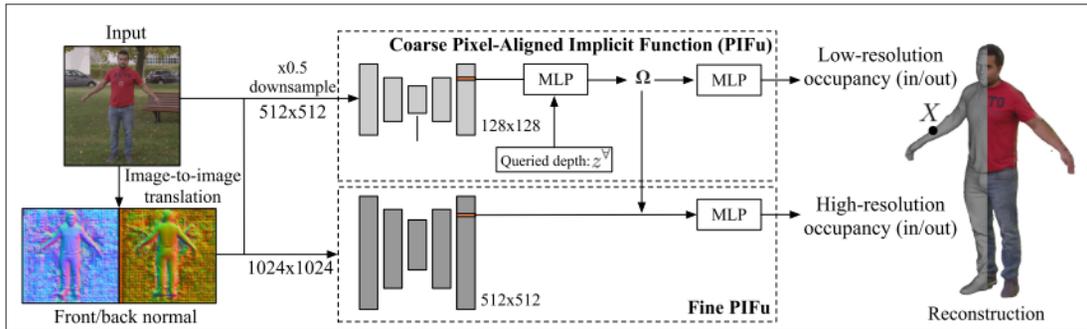
2) 단안 영상 기반의 휴먼 복원 기술

깊이 영상을 사용하지 않고 한 장의 영상만으로 전신 복원을 수행하는 기술은 기술의 난이도가 높지만 대신 휴대폰이나 웹캠 등으로 간편하게 아바타를 만들 수 있다는 장점이 있어 메타버스 기술의 대중화에 있어 가장 중요한 기술 중 하나라고 할 수 있다. 관련된 대표적 연구로는 BodyNet[9], PIFu[10], PIFuHD[11], Moulding Human[15] 등이 있으며 기술의 접근 방법과 특징은 다

음과 같다.

BodyNet[9]에서는 종단(end-to-end)간 학습이 가능한 휴먼 모델링 기술을 제안하였으며 입력 영상으로부터 2D/3D 자세 정보와 전경 정보를 추정된 뒤 이를 동시에 활용하여 볼륨 공간(implicit volume)을 예측하는 방법을 제안하였다. 하지만 종단 간 학습을 통해 볼륨 공간을 예측하기 위해서는 학습 시 많은 양의 메모리를 필요로 하여 높은 해상도의 복원이 어렵다. 높은 해상도의 모델 복원을 위해 PIFu[10]와 PIFuHD[11]에서는 볼륨 공간 전체를 예측하지 않고 개별 복셀(볼륨 공간을 구성하는 기본 단위)에 대해 점유(occupancy) 여부를 예측하는 접근 방법을 제안하였다. 다시 말해 학습되는 네트워크는 영상으로부터 추출된 특징 벡터를 입력으로 받아 특정 복셀이 객체의 내/외부에 존재하는지 여부를 예측하는데, 복셀 단위로 예측이 이루어지기 때문에 네트워크 구조가 단순하고 적은 양의 데이터만을 이용하여 학습하는 것이 가능하다는 장점이 있다.

결과적으로 PIFuHD의 경우 단일 영상을 사용하여 <그림 5>와 같은 수준의 복원이 가능함을 보였고, 최



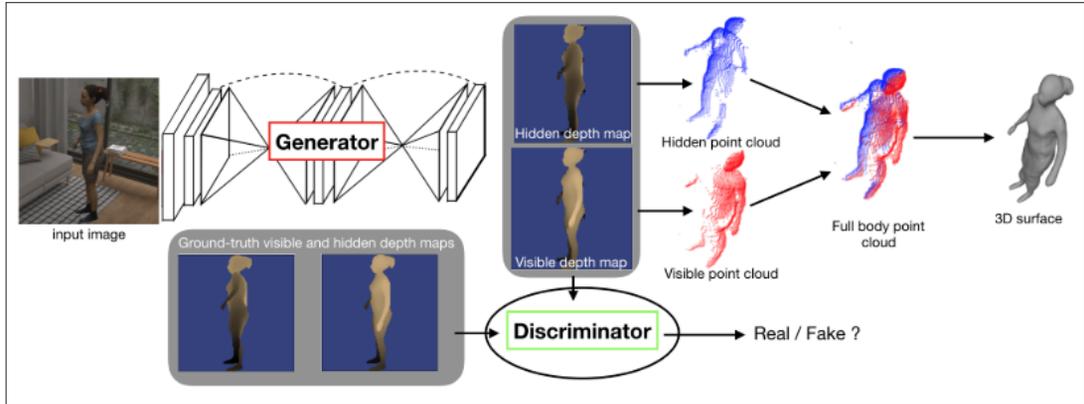
<그림 4> PIFuHD[11]: 단일 고해상도 영상에 대한 휴먼 복원 기술

근 수행되고 있는 많은 연구들이 본인들의 연구 결과를 PIFuHD와 비교하거나 발전시키는 형태로 이루어지고 있다. 예를 들어, PIFu, PIFuHD의 경우 복셀 단위로 예측을 하다보니 전체 형상(global shape)에 대한 제약사항이 없어 형상이 알 수 없는 형태로 일그러지는 경우가 있는데, Geo-PIFu[12]의 경우 형상 정보를 동시에 이용하여 이와 같은 문제점을 해결하였으며, StereoPIFu [13]의 경우 단안 영상을 양안으로 확장하여 안정성과 정확성을 한 단계 더 향상시키고자 하였다.

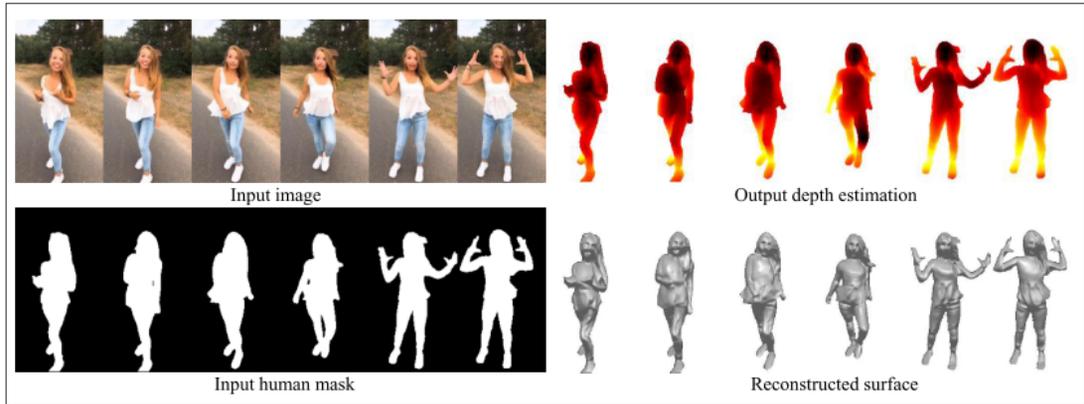
영상에서 볼륨 공간을 예측하는 것은 가장 널리 사용되는 접근 방법이지만 볼륨 공간 자체가 영상과 해상도가 동일한 3차원 공간이기 때문에 예측하는데 걸리는 시간이 오래 걸린다는 단점이 있고, 종단 간 학습 시 높은 메모리 사용량 때문에 고해상도 복원이 어렵다는 한계점이 있다. 이를 극복할 수 있는 접근 방법은 볼륨 공간 대신 여러 시점에 대한 깊이 영상을 생성하고 융합하는 방식이 있다. Moulding Humans[15]에서는 입력 영상에서 정면과 후면에 대한 깊이 영상을 예측



<그림 5> PIFuHD [11]: 단일 입력 영상에 대한 휴먼 복원 결과



<그림 6> Moulding Humans[15]: 단안 영상 기반의 3차원 모델 생성 과정



<그림 7> 소셜 미디어 동영상 기반의 깊이 영상 예측 기술[16]

하고 이를 융합하여 3차원 메쉬를 구성하는 적대적 신경망(adversarial neural network) 기반의 접근 방법을 제안하였다. 볼륨 공간을 예측하는 것과 대비했을 때 정확도는 높지 않지만 종단 간 학습을 통해 효과적으로 모델이 복원 가능성을 입증하였다.

최근에는 사람에 대한 깊이 영상을 추정하는 데 있어서 실측 모델이나 깊이 정보 없이 동영상 정보만을 사용하여 깊이 영상 시퀀스를 예측하는 방법이 미국 미네소타 대학에서 제안되었으며, 새롭게 데이터셋을 제작하거나 획득하지 않고 기존 소셜 미디어 데이터에

서 <그림 7>처럼 3D 모델링이 가능함을 보여주었다.

3. 휴먼 템플릿 생성 및 변형 모델링 기술

사람에 대한 템플릿 모델을 생성하고 리깅 한 뒤, 모션 캡처로 획득한 움직임을 적용하고 렌더링하는 과정을 통해 콘텐츠를 생성하는 기술은 영화나 게임 산업에서 오랫동안 사용된 기술이다. 최근에는 RealityCapture[18]와 같은 상용 소프트웨어의 완성도가 높아짐에 따라 손쉽게 템플릿 모델을 획득하고 조작



<그림 8> 넷플릭스 영화 Army of the dead의 한 장면

하는 것이 가능해졌다. 예를 들어 넷플릭스 영화 ‘Army of the dead[3]에서는 군중에 대한 템플릿 모델을 DSLR 카메라를 이용하여 촬영 현장에서 획득하였으며, 이를 <그림 8>과 같이 가상 군중 생성에 활용하였다.

메타버스 관점에서 바라봤을 때, 일반 사용자에 대한 템플릿 모델을 생성하고 실제 사람과 함께 움직이거나 다른 사람의 움직임을 모방하는 기술은 활용도가 매우 높은 기술이며 특히 광학식 마커를 사용하지 않고 전문가의 개입 없이 실제 사람처럼 움직이고 행동하는 기술은 고난이도의 기술이다. 핵심 기술로는 3차원 모델을 변형 가능한 템플릿 모델로 변형하는 기술, 이를 위

해 뼈대를 심고 뼈대와 정점 간의 연결성을 부여하는 과정이 필요한데 이를 리깅(rigging)과 스킨닝(skinning) 과정이라 한다. 그리고 템플릿 모델을 움직이기 위해서는 움직임 정보가 필요한데, 접근성 높은 움직임 획득을 위해서는 <그림 9>와 같이 순수 영상만을 이용한 3차원 자세 정보를 획득하는 과정이 필요하다. 마지막으로 임의의 움직임을 템플릿 모델에 적용하여 변형하는 단계가 필요하며 이때 목표로 하는 템플릿 모델의 체형이나 인체 구조에 맞게 움직임 정보를 맞춤제작하는 과정을 필요로 하게 된다. 최근에는 보다 자연스러운 움직임 생성을 위해 단순 스켈레톤 모델이 아닌 골격근을 함께



<그림 9> 영상 기반의 모션 캡처 스튜디오 솔루션을 제공하는 Capture사



<그림 10> 미국 Pinscreen사[5]의 가상 아바타 기술

모델링하는 방법론[19]이 제안되고 있으며, 앞으로 개발될 기술은 주요 관절 정보와 함께 손가락, 근육, 머리 카락, 얼굴 등 미세한 움직임 표현이 가능한 형태로 발전될 것이라 기대된다.

4. 뉴럴 휴먼 렌더링 기술

복원이 정확하게 이루어 지더라도 결국 사람의 눈에 VR 장치나 디스플레이를 통해 2D 영상 형태로 전달되는 경우가 대부분이기 때문에 마지막 단계에서 사실성 높은 영상을 생성하는 과정도 매우 중요하다. 영상 생성 기술은 적대적 신경망 기술이 대중화된 후 매우 빠르게 발전하고 있으며, 메타버스 관점에서는 가상 시점에 대한 영상을 생성하거나 합성하는 기술 [21], 합성된 영상의 사실감을 높여서 실제와 같이 느껴지게 해 주는 기술, 헤어스타일이나 의상을 변형 하는 기술 등이 이에 해당된다. <그림 10>은 미국 Pinscreen의 아바타 렌더링 결과로 눈동자, 머리카락, 의상에 대한 변화까지 실감나게 표현 가능한 형태로 렌더링이 가능하며 앞서 설명한 변형 모델링 기술을 적용하여 사실감 있는 콘텐츠 제작이 가능하다.

III. 맺음말

지금까지 미래 메타버스 시대의 핵심 기술인 휴먼 모델링 기술의 다양한 접근 방법과 함께 특징에 대해 살펴 보았다. 팬데믹으로 인한 비대면 산업의 급격한 성장으로 메타버스는 커다란 시대적 흐름이 되었고, 머지않아 많은 일상을 메타버스라는 공간 안에서 경험하는 시대가 올 것이라 예상이 된다. 아직은 자본력이 있는 콘텐츠 제작 기업에 의해 시장이 주도되고 있지만 그것은 어디까지나 메타버스 초창기의 형태일 것이고, 기술이 발전함에 따라 보다 많은 사람이 쉽고 편리하게 자신의 아바타를 생성하고 아바타를 통해 메타버스 공간에서 다른 사람과 소통하는 형태로 기술이 발전할 것이다. 끝으로 메타버스 시대를 맞이하기 위해서는 우리 모두가 메타버스가 갖는 역할과 의미, 그리고 가상의 아바타가 지능을 갖게 되었을 때 발생할 수 있는 사회적, 도덕적 문제와 같은 문제를 공론화하고 사회적 담론을 형성하려는 노력을 해야하지 않을까 조심스럽게 생각하며 본고를 마치고자 한다.

참고 문헌

- [1] LG유플러스, 세계 최초 4K 화질 3D AR 콘텐츠 제작, <http://www.uplus.co.kr>, 2019.
- [2] SKT 점프 스튜디오, 혼합현실 기술로 한층 더 진화하는 K팝 열풍, <https://www.jumpstudio.co.kr/news-detail?newsSeq=11>, 2021.
- [3] Wave XR, <https://wavexr.com/>
- [4] NETFLIX, Army of the dead, <https://www.netflix.com/kr/title/81046394>
- [5] Pinscreen, <https://www.pinscreen.com/>
- [6] RealityCapture, <https://www.capturingreality.com>
- [7] 조슈아 산토소, 전성호, 장주용, 박인규, 영상과 비디오로부터 3차원 휴먼 자세 및 형상 복원 기술, 한국방송-미디어공학회, Volume 26 Issue 3, 2021.
- [8] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan, High-Quality Streamable Free-Viewpoint Video, in SIGGRAPH, 2015.
- [9] G. Varol, D. Ceylan, B. Russell, J. Yang, E. Yumer, I. Laptev, and C. Schmid, **BodyNet**: Volumetric Inference of 3D Body Shapes, in ECCV, 2018.
- [10] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and Hao Li, **PIFu**: Pixel-aligned implicit function for high-resolution clothed human digitization, In Proceedings of the International Conference on Computer Vision (ICCV), 2019.
- [11] S. Saito, T. Simon, J. Saragih, and H. Joo, **PIFuHD**: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [12] T. He, J. Collomosse, H. Jin, and S. Soatto, **Geo-PIFu**: Geometry and Pixel Aligned Implicit Functions for Single-view Human Reconstruction, in NIPS, 2020.
- [13] Y. Hong, J. Zhang, B. Jiang, Y. Guo, L. Liu, and H. Bao, **StereoPIFu**: Depth Aware Clothed Human Digitization via Stereo Vision, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [14] Y. Lizhen, Z. Xiaochen, Y. Tao, W. Songtao, and L. Yebin, Z, **NormalGAN**: Learning Detailed 3D Human from a Single RGB-D Image, in ECCV, 2020.
- [15] V. Gabeur, J.-S. Franco, X. Martin, C. Schmid, and G. Rogez, **Moulding Humans**: Non-parametric 3D Human Shape Estimation from a Single Image”, in ICCV, 2019.
- [16] Y. Jafarian and H. S. Park, Learning High Fidelity Depths of Dressed Humans by Watching Social Media Dance Videos, in CVPR, 2021.
- [17] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson, and S. R. Fanello, **Fusion4D**: Real-time Performance Capture of Challenging Scenes, in ACM Trans. on Graphics (ToG), 2016.
- [18] R. A. Newcombe, D. Fox, and S. M. Seitz, **DynamicFusion**: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, in CVPR, 2015.
- [19] S. Lee, M. Park, K. Lee, and J. Lee, Scalable **Muscle-Actuated Human Simulation** and Control, in ACM Trans. on Graphics (ToG), 2019.
- [20] H. Joo, T. Simon, and Y. Sheikh, Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies, in CVPR, 2018.
- [21] R. M.-Brualla, P. Lincoln, A. Kowdle, C. Rhemann, D. B Goldman, C. Keskin, S. Seitz, S. Izadi, S. Fanello, R. Pandey, S. Yang, P. Pidyppenskiy, J. Taylor, J. Valentin, S. Khamis, P. Davidson, and A. Tkach, LookinGood: Enhancing Performance Capture with Real-time Neural Re-Rendering, in ACM Trans. on Graphics (ToG), 2018.
- [22] R. Kellnhofer, L. C. Jebe, A. Jones, R. Spicer, K. Pulli, and G. Wetzstein, Neural Lumigraph Rendering, in CVPR, 2021.
- [23] Z. Zheng, T. Yu, Y. Wei, Q. Dai, and Y. Liu, **DeepHuman**: 3D Human Reconstruction from a Single Image, in ICCV, 2019.
- [24] Z. Zheng, T. Yu, H. Li, K. Guo, Q. Dai, L. Fang, and Y. Liu, **HybridFusion**: Real-Time Performance Capture using a Single Depth Sensor and Sparse IMUs, in ECCV, 2018.

- [25] T. Yu, J. Zhao, Z. Zheng, K. Guo, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu, **DoubleFusion**: Real-Time Capture of Human Performance with Inner Body Shapes from a Single Depth Sensor, in Trans. on PAMI, 2020.
- [26] T. Yu, K. Guo, F. Xu, Y. Dong, Z. Su, J. Zhao, J. Li, Q. Dai, and Y. Liu, **BodyFusion**: Real-time Capture of Human Motion and Surface Geometry using a Single Depth Camera, in ICCV, 2017.
- [27] 8i Studio, <https://8i.com/>
- [28] Intel Studios, <https://newsroom.intel.com/press-kits/intel-studios/>
- [29] Canon Volumetric Video Studio, <https://global.canon/en/ws/>
- [30] Intel RealSense, <https://www.intelrealsense.com/>
- [31] Microsoft Kinect, <https://azure.microsoft.com/ko-kr/services/kinect-dk/>
- [32] StereoLabs, <https://www.stereolabs.com/zed/>

필자소개

박민규



- 2009년 : 건국대학교 학사
- 2011년 : 광주과학기술원 석사
- 2017년 : 광주과학기술원 박사
- 2016년 ~ 현재 : 한국전자기술연구원 선임연구원
- 주관심분야 : 3D 비전 및 휴먼 모델링

강주미



- 2015년 : 한국교통대학교 학사
- 2017년 : 한양대학교 석사
- 2017년 ~ 현재 : 한국전자기술연구원 연구원
- 주관심분야 : 컴퓨터비전 및 색채 보정

윤주홍



- 2008년 : 성균관대학교 학사
- 2009년 : 광주과학기술원 석사
- 2014년 : 광주과학기술원 박사
- 2014년 ~ 현재 : 한국전자기술연구원 선임연구원
- 주관심분야 : 컴퓨터비전 및 휴먼 변형 모델링