

특집논문 (Special Paper)

방송공학회논문지 제27권 제5호, 2022년 9월 (JBE Vol.27, No.5, September 2022)

<https://doi.org/10.5909/JBE.2022.27.5.751>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

극사실적 메타휴먼을 위한 3D 볼류메트릭 캡처 기반의 동적 페이스 제작

오문석^{a)}, 한규훈^{b)}, 서영호^{c)*}

3D Volumetric Capture-based Dynamic Face Production for Hyper-Realistic Metahuman

Moon-Seok Oh^{b)}, Gyu-Hoon Han^{b)}, and Young-Ho Seo^{c)*}

요 약

디지털 기술의 발전에 따라 메타버스가 콘텐츠 시장의 주요 트렌드로 자리하면서 고품질의 3D(dimension) 모델을 생성하는 기술에 대한 수요가 급증하고 있다. 이에 따라 디지털 휴먼으로 대표되는 고품질 3D 가상 인간의 제작과 관련된 다양한 기술적 시도가 이루어지고 있다. 3D 볼류메트릭 캡처는 기존의 3D 모델 생성 방식보다 빠르고 정밀한 3D 인체모형을 생성할 수 있는 기술로 각광 받고 있다. 본 연구에서는 볼류메트릭 3D 및 4D 모델 생성 분야에서 적용되는 기술들과 콘텐츠 제작의 애로사항에 대해 실제 사례를 바탕으로 3D 고정밀 페이스 제작 기술의 분석을 시도한다. 그리고 실제 볼류메트릭 캡처를 통한 모델의 구현 사례를 바탕으로 3D 가상 인간의 얼굴 제작에 대한 기술들을 고찰하고 효율적인 휴먼 페이스 생성을 위한 그래픽스 파이프라인을 활용하여 새로운 메타휴먼을 제작하였다.

Abstract

With the development of digital graphics technology, the metaverse has become a significant trend in the content market. The demand for technology that generates high-quality 3D (dimension) models is rapidly increasing. Accordingly, various technical attempts are being made to create high-quality 3D virtual humans represented by digital humans. 3D volumetric capture is spotlighted as a technology that can create a 3D manikin faster and more precisely than the existing 3D model creation method. In this study, we try to analyze 3D high-precision facial production technology based on practical cases of the difficulties in content production and technologies applied in volumetric 3D and 4D model creation. Based on the actual model implementation case through 3D volumetric capture, we considered techniques for 3D virtual human face production and produced a new metahuman using a graphics pipeline for an efficient human facial generation.

Keyword : hyper-realistic, metahuman, volumetric, face model, 3d scanning, metaverse

1. 서론

1. 연구의 배경

“수아”, “로지”, “릴 미켈라”, “이마” 등 국적이 무엇인지 알기 힘든 이름들이다. 이들은 현재 국내의 온라인을 중심으로 활동하고 있는 가상인간 즉, 디지털 휴먼, 혹은 메타휴먼들이다. 미국 스타트업 ‘브러드’가 만든 “릴 미켈라”의 경우 광고모델로 활동하며 2021년도 약 120억원 이상의 수익을 올렸다^[1]. 우리나라의 경우에는 “로지”를 통해 만들어진 가상인간이 광고모델로써 충분한 역할과 수익모델이 될 수 있다는 것을 검증하였다^[2]. 시간과 공간의 제약이 없고, 스캔들 및 사고의 우려도 없는 안전성과 지적재산권의 자유로움은 이들의 활용이 더욱 증가될 수 있을 것이라는 것에 무게를 실어주고 있다. 뿐만 아니라 최근에는 인공지능 기술의 발달을 통해 인공지능과 가상인간이 결합하면서 가상인간에 TTS (text-to-speech)^[3]와 STF (speech-to-facial)^[4] 기술을 접목하여 스스로 말하고 표정 지을 수 있는 기능을 부여하고 있는 중이다. 따라서 최근에 콘텐츠 및 그래픽스와 관련된 업계에서는 3D 가상인간 혹은 메타휴먼의 관심이 매우 높다^[5]. 최근 급격히 발전한 그래픽스 엔진 기술과 GPU(general processing unit)의 눈부신 발전으로 인해 가상인간의 품질과 제작 환경은 하루가 다르게 향상되고 있다. 또한 여기에 모션 캡처 기술의 발전하면서 가상인간에게 쉽게 움직임을 부여할 수 있는 환경도 조성되었다. 온마인드가 “수아”를 통해 구현한 기술의 핵심은 실시간 모션

캡처를 바탕으로 “수아”를 실시간으로 자연스럽게 움직일 수 있는 기술이다. ‘유니티’의 게임엔진에 가상인간의 능력을 결합하여 실사와 함께 공존할 수 있는 3차원 공간 속에서 가상인간이 실시간으로 움직이는 기술을 구현했다. 기술마다 장단점이 있기 때문에 같은 가상인간도 목적과 용도에 따라 제작 회사에 따라 자신들만의 방법으로 제작된다. 실시간 동작이 필요한 가상인간이라면 게임엔진 기반의 3D 기술을 활용하고, 고화질영상 제작 및 소셜네트워크용으로 업로드할 영상 등의 경우에는 매우 정교하게 그래픽 편집 도구를 이용하여 제작한다. 일반적으로 하나의 가상인간에 대해서 몇 가지 모델을 만들어두고 이들을 활용 용도에 따라서 적당하게 편집 및 가공하여 사용한다.

가상인간 분야에서 포토리얼리스틱 가상인간 분야의 가장 중요한 점은 얼마나 사람과 똑같이 제작할 수 있는가이다. 소위 말하는 ‘불쾌한 골짜기’를 완전히 벗어나서 실제 인간과 차이를 잘 느끼지 못하는 영역까지 나아갈 수 있는 기술까지 넘볼 수 있는 시대가 도래하고 있다. ‘불쾌한 골짜기’ 이론은 로봇이나 가상인간이 점점 사람의 모습에 가까워질수록 호감도가 상승하다가 일정 수준을 넘어서면 오히려 거부감이 강해진다는 이론이다^[6]. 하지만 이 수준을 뛰어넘어 인간과 거의 구별이 불가능한 수준이 되면 다시 호감도가 상승할 수 있다. 이와 같은 경험은 현재 많은 영화들에서 빈번하게 이루어지고 있다. 많은 SF 영화는 거의 모든 장면과 인물들이 오랜 비용과 인력을 투입하여 만들어진 것이라는 것을 우리는 알고 있지만 아무런 거부감 없이 실제와 같은 것으로 착각하면서 자연스럽게 받아들이고 있다. 실제 인간과 동일한 실제감과 정밀함을 갖는 가상인간을 제작하는 데 가장 힘든 영역으로는 얼굴과 표정에 대한 표현이다. 얼굴의 표정을 표현하기 위해 얼굴을 정교하게 리깅하고, 다양한 얼굴 표정들 사이의 변환하기 위한 모핑을 동시에 사용하는 방법이 일반적으로 활용된다. 얼굴 근육의 동작을 적게는 수십 개, 많게는 수백 개의 패턴으로 만들어 놓고 지어야 하는 표정에 따라서 조합해야 한다. 즉, 얼굴의 섬세한 표정과 감정에 따른 근육들의 움직임을 정확하게 표현하기 위해서는 얼굴의 해부학적 구조에 대한 지식을 바탕으로 그래픽 도구를 이용한 재현 기술이 발전되어야 한다^{[7][8]}.

a) 광운대학교 미디어커뮤니케이션학부(Kwangwoon University School of Media & Communication)

b) 오모션 주식회사(5Motion.Inc)

c) 광운대학교 전자재료공학과(Dept. of Materials Eng., Kwnagwoon University)

‡ Corresponding Author : 서영호(Young-Ho Seo)

E-mail: yhseo@kw.ac.kr

Tel: +82-2-940-8362

ORCID: <https://orcid.org/0000-0003-1046-395X>

※This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2022. The present Research has been conducted by the Research Grant of Kwangwoon University in 2022.

· Manuscript August 9, 2022; Revised August 22, 2022; Accepted August 22, 2022.

2. 연구목적 및 방법

실제 사람과 구별이 어려울 정도의 실감을 제공하는 가상인간에 인공지능이 결합된다면 가장 먼저 아나운서나 캐스터, 그리고 쇼호스트 등의 역할을 수행하는 것이 가능해진다. 이러한 직업들이 가상인간에 의해 가장 위협받을 수 있는 이유는 거의 유사한 포맷으로 거의 유사한 내용을 며칠 단위로 반복하기 때문이다. 이러한 작업들을 자동화시키고 가상인간에게 맡긴다면 매우 효율적인 업무처리가 가능할 것이다. 가상인간은 사람이 지정해놓은 동작과 움직임을 자연스럽게 표현할 수 있다. 이러한 기능을 활용한다면 일상에서도 온라인을 기반으로 하는 운동 코칭이나 각종 교육 및 훈련의 강사로서 가상인간이 대신하는 것은 물론이고, 고객을 응대하거나 매우 친절한 자세로 한결같이 고객 상담도 가능할 것이다. 이와 같은 다양한 행위들은 디지털 트윈일 경우에 더욱 장점을 가질 수 있다. 디지털 트윈은 실제 사람을 대상으로 하여 3D 볼류메트릭 스캐닝을 하고, 대상 인물의 움직임과 표정 등을 동일하게 표현할 있도록 리깅 및 모핑이 가능한 가상인간의 한 형태이다. 디지털 트윈과 더불어 최근에 많은 발전을 이룬 분야는 이른바 “딥페이크”라고 불리는 기술이다. 딥페이크는 특정인을 그대로 재현해내는 기술을 포함하여 다양한 인물을 합성하여 새로운 인물로 재현해내는 기술까지 포함된다. 딥페이크 기술은 최근에 발전한 인공지능 중에서 생성형 딥러닝 모델을 기반으로 하고 있다. 이와 같은 기술들은 3D 그래픽스 기술

과 만나면서 그 영역을 3차원과 4차원 영역으로 확장하고 있고, 새롭게 개발되고 있는 다양한 3D 그래픽스 기술들을 장착하여 이전보다 훨씬 정교하고 실제와 같은 결과물들을 내놓고 있다. 그러나 아직까지 딥러닝 기술을 이용한 3D 그래픽스 모델 생성 결과물들은 고품질이 아니기 때문에 콘텐츠로 활용하기에는 매우 제한적이다. 따라서 아직까지는 많은 부분 수작업을 통해서 3D 그래픽스 형태의 가상인간을 제작하고 있다. 가상인간을 제작하는 많은 기업들은 그들만의 다양한 노하우와 자체적인 제작방식을 이용하고 있기 때문에 가상인간을 제작하는 표준화된 방법이라는 것은 존재하지 않고, 제작사별로 가상인간을 제작하는데 소요되는 비용도 매우 다양하다.

우리는 본 논문을 통해서 이와 같은 가상인간을 제작하는데 사용되는 다양한 3D 그래픽스 기술들 중에서 본 연구팀이 보유한 기술들을 이용하여 실제 콘텐츠를 제작할 수 있는 수준의 3D 페이스 모델을 제작한 사례를 보이고자 한다. 본 연구팀이 제작하고자 하는 디지털트윈 방식의 가상인간은 방송 및 영화에서 사용될 수 있는 수준의 품질을 목표로 하고 있고, 가상인간의 제작을 위한 바디와 페이스 등은 모두 3D 볼류메트릭 캡처를 통해서 자동화하여 생성하는 것을 목표로 하고, 특히 본 논문에서는 페이스 생성에 대해 중점적으로 다루고자 한다. 이러한 과정을 바탕으로 실사 기반의 3D 페이스 모델의 제작 방법에 대한 한 가지 파이프라인을 설계해보고, 향후 3D 페이스 모델의 제작에 있어서 발전 방향 등에 대해 제시해 보고자 한다.

표 1. 디지털 휴먼과 관련된 정의

Table 1. Relative definition of digital human

Researcher	Definition
Bo-eun Kwak, Jeong-yoon Heo (곽보은·허정윤)(2021) ^[9]	3D virtual human that looks and speaks like a real person
Soo-ho Kang, Mi-ae Son (강수호·손미애)(2012) ^[10]	Objects that can imitate body features, postures, and motions of people performing specific tasks
Se-young Kim, Jeong-yoon Heo (김세영·허정윤)(2021) ^[11]	A realistically generated 3D human model in terms of movement of a form developed from the concept of an avatar.
Seunghwan Lee, Sangyeol Han (이승환·한상열)(2021) ^[12]	3D virtual human with a shape similar to real human appearance and behavior

II. 가상인간 제작의 이론적 배경

1. 디지털 휴먼과 메타휴먼

디지털 휴먼의 정의는 디지털 휴먼이 적용되는 분야와 연구 목적에 따라 다양하게 나타난다. 우리는 다양한 연구자들에 의해 정의된 디지털 휴먼의 의미에 대해 표 1에 정리하였다. 가장 보편적인 용례에 부합하는 정의로서는 “실제 사람과 같은 외형과 말을 하는 3D 가상인간”이며, 메타버스 콘텐츠의 측면에서 고려할 때는 “기존 아바타의 개념에서 발전된 형태 및 움직임의 측면에서 사실적인 결과물

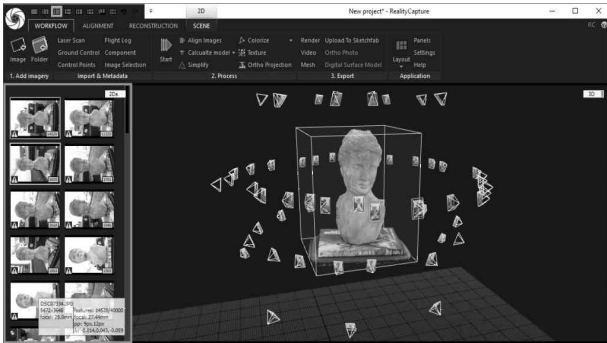


그림 1. 포토그래메트리를 통한 3D 모델링 개요 (출처 : CapturingReality)
Fig. 1. 3D modeling by the photogrammetry

“생성하는 3D 모델”이라고 정의할 수 있다. 다양한 연구 사례에서 디지털 휴먼을 정의한 내용을 확인했을 때 디지털 휴먼의 공통적인 필수 조건은 ①사람의 외형을 모사한 정밀한 인체 묘사 ②실제 사람의 행동과 같은 사실적인 움직임 구현 ③가상공간에서 사람의 역할을 대체하는 역할이라고 할 수 있다^[9].

점진적인 컴퓨터그래픽스 기술의 발전으로 인해 디지털 휴먼은 더욱 정밀하고 정교하게 표현되기 시작했다. 또한 그동안 축적된 빅데이터와 인공지능의 도움을 통해 아카이브화 된 3D 디지털 자산(3D Digital Asset)을 구축하고 이를 활용하여 빠르고 간편하게 디지털 휴먼을 만드는 기술이 등장하였다. 2021년, 미국의 Epic Games社는 자사의 게임엔진인 언리얼의 실시간 렌더링 기능과 기존의 3D Asset을 활용하여 쉽고 빠르게 디지털 휴먼을 제작할 수 있는 툴인 “메타휴먼 크리에이터(Metahuman Creator)”를 공개하였고 이를 통해 제작되는 디지털 휴먼에 대해 “메타휴먼”이라고 명명하였다. 메타휴먼은 표 2와 같이 디지털 휴먼과

구분되는 발전된 요소들과 특징을 가진다^[10].

메타휴먼의 가장 큰 특징은 3D Asset을 통해 빠르게 제작할 수 있다는 점이며, 제작에 사용되는 3D Asset은 사전에 리깅(Rigging) 3D 모델에 움직임 구현을 위한 뼈대 및 관절을 삽입하는 제작상의 과정이 완료되어 자유로운 움직임을 구현할 수 있다는 점이다. 이러한 메타휴먼의 특징에 대해 오문석·서영호는 메타휴먼을 기존의 디지털 휴먼의 제작 방식상의 단점을 보완한 “포스트 디지털 휴먼”으로 규정하였으며, 메타휴먼을 “능동적 가상인간”으로 정의하였다. 또한 3D Asset을 활용하여 제작하거나 볼류메트릭 캡처 등을 통해 제작되고 모션캡처 또는 모션캡처 데이터를 적용하여 움직임을 구현할 수 있는 디지털 휴먼을 지칭하는 용어로서 메타휴먼을 사용하였다^[11].

2. 포토그래메트리

포토그래메트리(Photogrammetry)는 다방면에서 촬영된 복수의 이미지를 바탕으로 3D 입체 모델을 구현하는 기술 중 하나이다. 획득한 이미지를 사용해 포토그래메트리를 통한 3D 모델링은 그림 1과 같이 이루어진다^[12].

포토그래메트리의 장점은 디지털카메라와 같이 이미지를 획득할 수 있는 도구만 있으면 별도의 장비 없이 3D 스캐너를 통한 결과물과 유사한 수준의 3D 모델링 결과물을 얻을 수 있다는 점이다. 또한 고화질로 촬영된 이미지 기반의 합성 데이터를 통해 이미지에서 추출한 고품질의 텍스처를 적용하여 극사실적인 표현이 가능한 장점이 있다. 그 외에도 동일한 조건을 유지한 상태에서 다각도에서 촬영된 이미지만 있으면 촬영 대상의 크기나 면적에 구애받지 않

표 2. 메타휴먼과 디지털 휴먼의 비교

Table 2. Comparison of metahuman and conventional digital human

Metahuman	Main Items	Digital Human
Production using 3D Asset	Production Method	Sculpting or utilizing existing 3D assets
Asset Selection and Parameter Manipulation by GUI	3D Model Modification	Modification of mesh data and topology
3D Asset with pre-rigging completed Implementation of animation through application of keyframes and motion data	Movement Implementation	Separate Rigging Tasks Animation via keyframes
Metahuman Creator Unreal Engine	Production Tools	3D Production tools such as 3Ds Max, Maya, Cinema 4D, etc.

고 3D로 구현할 수 있다는 장점 때문에 건물의 입체 모형을 제작하거나 드론을 통한 촬영 데이터를 바탕으로 3D 입체 지도를 제작하는 등 다양한 분야에서 활용되고 있다.

이전까지의 포토그래메트리는 고화질의 디지털카메라와 고성능 컴퓨터를 통한 이미지 합성 방식을 사용하였지만, 최근에는 클라우드 컴퓨팅 및 스마트폰의 성능 발전으로 인해 스마트폰으로 이미지를 획득하고 3D 객체를 제작할 수 있는 기술이 등장하였다. 2022년 4월, CapturingReality社は 자사가 제공하고 있는 포토그래메트리 서비스인 Reality Capture의 모바일 버전인 Reality Scan을 발표하였다. 이를 통해 스마트폰으로 획득한 이미지를 서버로 전송하여 합성하고, 합성된 결과를 실시간으로 확인 및 출력하여 3D 객체로 활용할 수 있게 되었다. 이를 통해 기존의 고가의 3D 스캐너를 대신하여 쉽고 빠르게 3D 객체를 생산할 수 있는 환경이 열리게 될 것으로 예상된다. 그림 2는 Reality Scan의 작동 방식을 보여준다.



그림 2. 스마트폰을 활용한 포토그래메트리 예시 (출처 : CapturingReality)
Fig. 2. Example of the photogrammetry using a smartphone

포토그래메트리는 그 과정상 3D 스캐닝과 유사한 방식으로 진행되나 결과물에 측면에서는 3D 스캐닝보다 고품질의 결과물을 얻을 수 있다는 장점이 있다. 그러나 포토그래메트리로 얻을 수 있는 결과물은 단순 모델링 데이터로 움직임을 구현하기 위해서는 별도의 리깅 과정을 거쳐야 한다. 이미지 합성 및 처리기술의 발전으로 인해 현재 포토그래메트리는 빠르게 3D 스캐닝을 대체하는 기술로서 자리잡고 있다. Epic Games社의 경우 포토그래메트리 솔루션을 개발하는 CapturingReality社를 인수하여 해당 기술을

자사의 언리얼 엔진에 통합시켰으며 이를 통해 다양한 3D

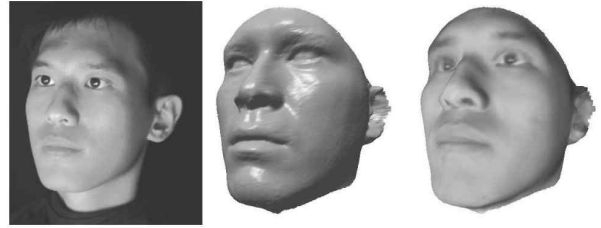


그림 3. 실루엣을 기반으로 한 3D 페이스 모델 제작 (출처 : Visual Computing Group)
Fig. 3. 3D face model making based on the silhouette processing

에셋을 제작하는데 활용하고 있다.

3. 페이스 캡처

일반적으로 사용되는 페이스 캡처는 복합적인 과정을 의미한다. 그 과정을 세분화하면 ① 얼굴 모형을 만드는 “페이스 모델 캡처” ② 3D 페이스 모델의 움직임 구현을 위한 “페이스 리깅” ③ 얼굴의 움직임을 추적하는 “페이스 모션 캡처”로 나눌 수 있다.

3.1 페이스 모델 캡처

페이스 모델 캡처는 사람의 얼굴 모형을 3D로 제작하는 전반적인 과정을 말한다. 이 과정에는 다양한 방식이 있으나 일반적으로 ① 스컬핑을 통한 제작 ② 이미지를 기반으로 한 3D 모델 생성 ③ 3D 스캔 또는 포토그래메트리를 통한 제작방식이 사용된다. 페이스 모델 캡처는 3D로 사람의 얼굴을 제작하는 것을 주목적으로 하며 최대한 정교한 외형을 구현하는 방식으로 진행된다. 제작 시 3D 모델링의 움직임 등은 고려하지 않고 제작되는 만큼 최대한 디테일한 표현에 집중하는 경향이 있다. 따라서 페이스 모델 캡처를 통해 제작된 3D 페이스 모델의 경우 비교적 정밀도가 높은 특징을 가진다. 그림 3은 이미지를 기반으로 3D 페이스 모델을 캡처하는 방식의 예시이다.

3.2 페이스 리깅

페이스 리깅은 페이스 모델 캡처를 통해 제작된 3D 페이스 모델의 변형을 위해 얼굴 각 부분에 근육, 골격과 같이

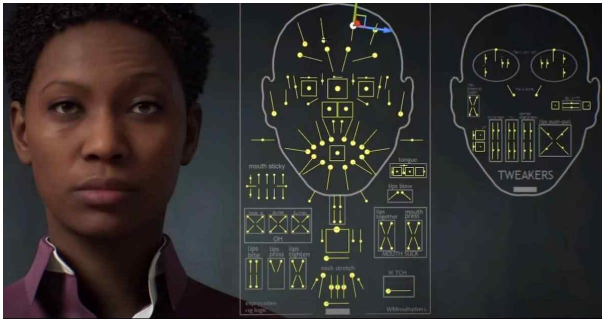


그림 4. 메타휴먼 크리에이터를 통한 페이스리깅 (출처 : Epic Games)
Fig. 4. Facial rigging using the metahuman creator

움직임을 구현할 수 있는 기준점(rig joint)을 지정하는 과정을 말한다. 페이스얼 캡처를 통해 제작된 3D 페이스 모델의 경우 3D 모델에 리깅이 적용되어 있지 않기 때문에 표정 구현 및 자연스러운 움직임을 구현하기 위해 움직이는 모든 얼굴 근육에 대응하는 폴리곤을 수동으로 조작해야 하며 모션 캡처 데이터의 적용이 불가능하다. 3D 페이스 모델에 리깅이 적용되면 페이스얼 모션 캡처를 통해 획득한 모션 데이터를 적용하여 실제 사람이 짓는 표정과 동일한 수준의 자연스러운 표정과 움직임을 구현할 수 있다. 그림 4는 메타휴먼 크리에이터에서 제공하는 페이스 리깅의 예시이다.

3.3 페이스얼 모션 캡처

페이스얼 모션 캡처는 이미지 또는 광학식 스캔 등을 통해 얼굴의 형태와 움직임을 추적하고 이를 통해 형태와 특징을 파악하여 움직임을 추적하고 이를 데이터화 하는 과정을 말한다. 페이스얼 모션 캡처를 위해서는 실시간으로 이미지를 획득 및 획득한 이미지를 분석하고 메쉬 데이터로 구현할 수 있는 기술을 필요로 한다.

페이스얼 모션 캡처는 3D 페이스 모델의 제작보다는 얼굴의 움직임을 구현하는데 주목적이 있다. 따라서 페이스얼 모션 캡처를 통해 획득한 데이터는 얼굴의 특징을 파악하여 자연스러운 얼굴의 변형(morphing)을 통한 움직임을 구현하는데 주로 활용된다. 그림 5는 페이스얼 모션 캡처의 예시이다.

페이스얼 모션 캡처는 사전에 모션 캡처를 통해 모션 데이터를 획득하고 3D 모델을 활용한 애니메이션을 제작할 때 별도로 메타 데이터를 적용하는 방식으로 많이 적용되나 최근에는 실시간 렌더러 기술의 발전으로 인해 실시간으로

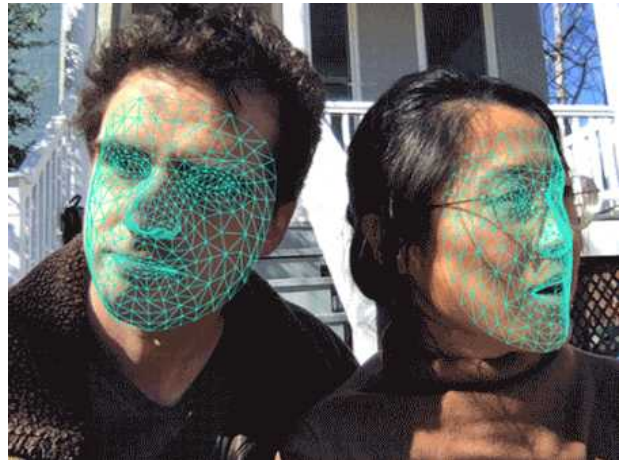


그림 5. 이미지 기반 페이스얼 모션 캡처의 예시 (출처 : mediapipe)
Fig. 5. Example of image-based facial motion capture

페이스얼 모션 캡처 데이터를 3D 페이스 모델에 반영하여 사용하는 추세이다.

3D 페이스 모델을 구현하기 위해서는 페이스얼 캡처와 관련된 세 가지 기술이 조화롭게 적용되어야 한다. 스켈핑 또는 포토그래메트리 등을 통해 정밀한 3D 모델을 제작하고, 제작된 모델에 리깅을 통해 움직임을 구현할 수 있도록 하며 여기에 정밀한 모션 캡처 데이터를 적용하는 것으로 실제와 같은 외형과 움직임을 보이는 페이스 모델을 제작할 수 있다. 이처럼 단순히 외형만을 본뜬 3D 모델이 아닌 변형이 가능한 페이스 모델을 3D Morphable Model이라고 부르며(B.Egger et al. 2020)^[13], 본 연구에서는 모션 캡처 등의 데이터를 통해 움직임에 대응할 수 있는 3D 모델을 페이스 모델과 구분하기 위해 3D 페이스얼 모델(3D Facial Model)로 정의하였다^[14].

본 연구에서는 국내 모션 캡처 전문기업인 모션테크놀로지사의 기술지원과 JTBC와의 협업을 통해 이수진 기자의 디지털 더블(Digital Double)을 제작하였다. 이 과정에서 볼류메트릭 캡처와 포토그래메트리 합성을 통해 기초가 되는 페이스 모델을 제작하고 후보정을 통해 정밀한 3D 페이스 모델을 제작하였다. 이후 제작된 페이스 모델에 리깅을 적용하고 언리얼 엔진을 활용하여 실시간 페이스얼 모션 캡처와 실시간 렌더링을 통해 실제 모델의 움직임이 실시간으로 반영되는 정밀한 페이스 모델을 구현하였다.

III. 실사 페이스의 제작

1. 전체 제작 과정

극사실적 메타휴먼의 페이스 제작을 위한 전체 처리과정은 그림 6과 같다. 이 방식을 통해 제작된 메타휴먼의 페이스는 실사 텍스처를 그대로 사용하고, 다양한 표정의 변화가 가능해야 한다. 또한 표정 변화는 다양한 페이스얼 모션 캡처 장비와 실시간 엔진 상에서 연동이 가능해야 한다. 다양한 표정을 취한 후에 각각의 표정을 획득하고, 이들을 다시점 카메라를 이용하여 각각 하나의 3D 모델로 합성한다. 여기에 미리 제작한 얼굴에 대한 3D 토폴로지를 매핑하여 리메싱을 수행하여 각 얼굴 표정별 3D 페이스 모델을 생성한다. 얼굴 표정의 개수 등은 응용 분야와 사람의 특성에

따라서 달라질 수 있다. 각 얼굴 표정들에 대한 3D 페이스 모델을 생성한 후에 페이스웨어 툴을 이용하여 모핑이 가능한 3D 페이스 모델을 생성한다.

2. 페이스 모델 캡처

정밀한 페이스 모델의 제작을 위한 베이스가 되는 페이스 모델을 제작하기 위해 가이드 모델을 대상으로 볼류메트릭 캡처를 활용하여 기본 데이터를 획득하였다. 40대의 카메라를 통해 전방위에서 고해상도의 스틸 이미지를 획득하고, 획득한 이미지를 포토그래메트리 방식으로 합성하는 방식으로 진행하였다. 볼류메트릭 캡처의 정밀도 향상 및 포토그래메트리 합성 시 정교한 표현이 어려운 머리카락은 후보정 단계에서 별도의 에셋을 통해 합성하기 위해 가발

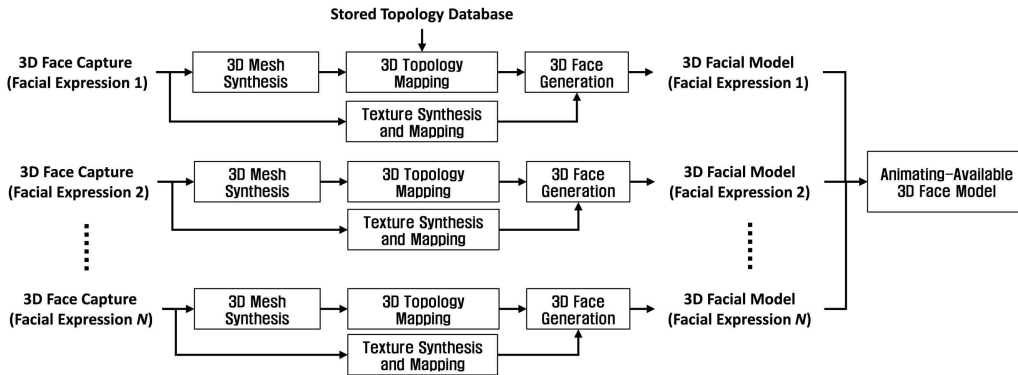


그림 6. 극사실적 3D 페이스 제작 과정
 Fig. 6. Hyper-realistic 3D face production process

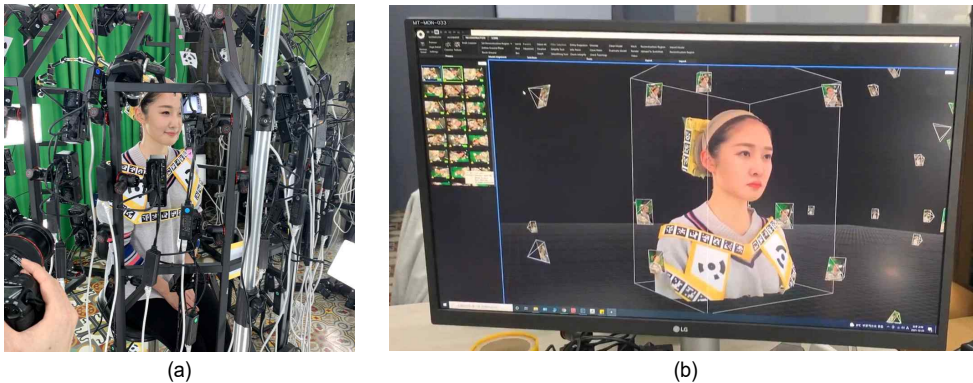


그림 7. 3D 볼류메트릭을 이용한 3D 페이스얼 캡처 과정 (a) 볼류메트릭 캡처, (b) 포토그래메트리 합성
 Fig. 7. 3D facial capture process using 3D volumetric (a) volumetric capture system, (b) photogrammetric synthesis

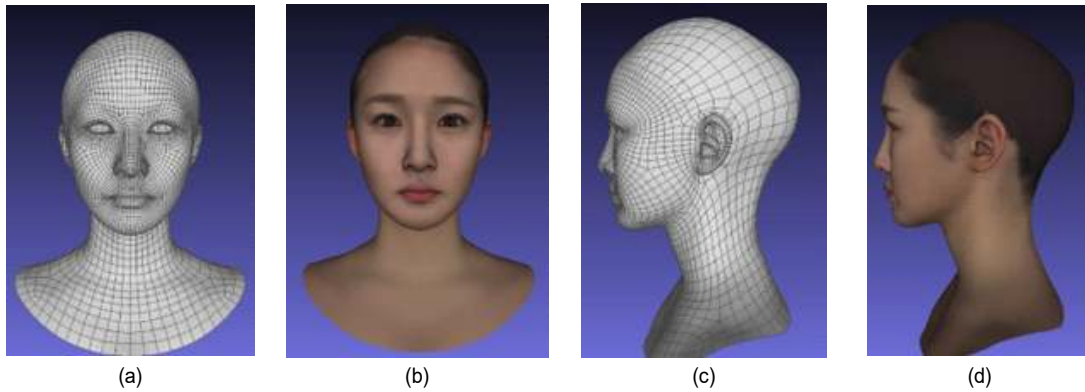


그림 8. 3D 볼류메트릭을 이용한 3D 페이스얼 캡처 결과
 Fig. 8. 3D facial capture result using 3D volumetric scanning

망을 착용하였다. 이 과정을 통해 페이스 모델의 기본 형태를 획득할 수 있다. 그림 7은 페이스 모델 제작과정이다.

포토그래메트리 합성을 통해 획득한 3D 모델은 이후 후보정을 통해 머리카락의 합성 및 얼굴 주요 부위의 보정 및 수정을 거치게 된다. 이를 위해 합성에 불필요한 부분을 제거하고 리메쉬(remesh) 및 리토폴로지(retopology)를 통해 기초적인 페이스 모델의 형태를 구성해주었다. 그림 8은 볼류메트릭 캡처와 포토그래메트리를 통해 획득한 페이스 모델을 기반으로 수정한 결과물이다.

3. 페이스얼 모델 후보정

볼류메트릭 캡처로 제작되고 수정을 거친 페이스 모델에 대해 후보정을 통해 페이스 모델의 디테일을 보강하기 위해 ① 에셋을 통한 헤어스타일 추가 ② 얼굴 셰이프 수정 ③ 텍스처 작업 및 스킨 셰이딩 등을 진행하였다. 또한, 후보정 작업에서는 변형과 움직임 구현이 불가능한 3D 모델을 리깅 및 움직임 적용이 가능하도록 구조를 수정하는 과정도 함께 진행하였다. 이 과정에서 리깅 적용을 위한 안구 삽입 및 구강구조를 구현하여 안구의 움직임과 발화를 표현할 수 있도록 페이스 모델을 수정하였다. 그림 9는 제작된 페이스 모델의 후가공 전과 후를 비교한 그림이다.

최종적으로 후보정을 통해 텍스처와 스킨셰이딩 등이 적용된 바디모델 결합된 3D 페이스 모델의 결과물은 그림 10과 같다.



그림 9. 후보정을 통한 페이스 모델 수정 (a) 후가공 전, (b) 후가공 후
 Fig. 9. Face model modification through post-production (a) before, (b) after

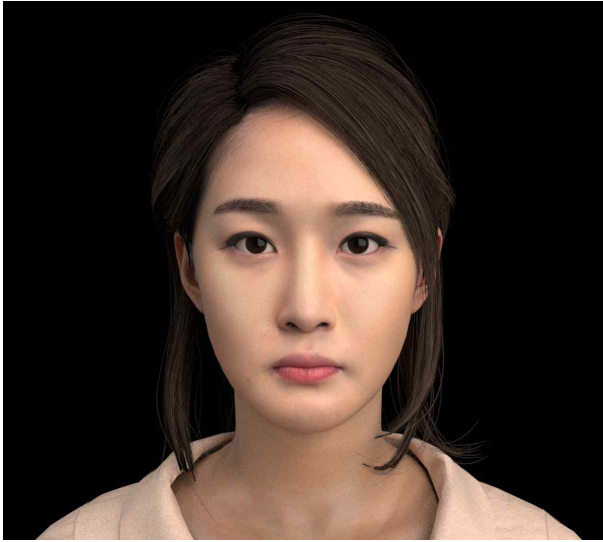


그림 10. 3D 볼류메트릭 스캐닝과 합성을 이용한 3D 페이스 제작 결과
Fig. 10. 3D face production result using 3D volumetric scanning and synthesis

4. 리깅 및 실시간 페이스 모션 캡처

후보정이 완료된 페이스 모델을 기반으로 페이스 리깅을 진행하였다. 페이스 리깅을 통해서 모션 캡처 데이터를 통해 구현될 시선의 움직임에 따른 안구의 움직임과 눈의 깜빡임, 발화 시 구강구조의 변화와 다양한 표정 변화에 대응할 수 있도록 하였다^[15]. 페이스 리깅은 Autodesk Maya를 사용하여 진행하였으며 모션 캡처 데이터의 적용 없이 파라미터의 변경을 통해 다양한 표정을 구현한 결과물은 그림 11과 같다.

페이스 모션 캡처는 Apple ARKit을 통해 언리얼과 연동되는 Live Link Face 애플리케이션을 사용하여 실시간으로 모션 캡처 데이터를 획득하였다. 획득된 데이터는 언리얼 엔진에 셋업된 3D 페이스 모델에 실시간으로 반영되며 언리얼 엔진을 통해 실시간 렌더링으로 그 결과물을 확인할 수 있었다. 그림 11은 아이폰과 Live Link Face 애플리케이션을 통해 실시간 페이스 모션 캡처를 진행하는 과정과 실시간으로 출력된 그 결과물이다.

3D 페이스 모델을 제작하는데 있어서 가장 까다로운 작업은 자연스러운 얼굴의 표정을 만들어낼 수 있는 페이스 리깅 작업이다. 바디의 리깅과 달리 페이스 리깅은 그 사람

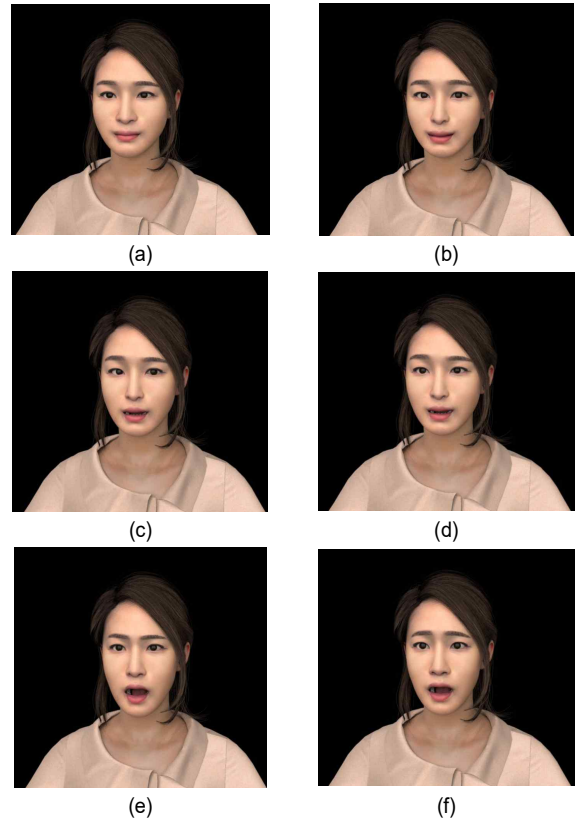


그림 11. 리깅을 통해 구현된 다양한 표정
Fig. 11. Various facial expression by facial motion capture

의 음성에 따른 표정을 잘 담아낼 수 있어야 하기 때문에 수작업으로 수행하는 것보다는 자동화된 프로세스를 따르는 것이 훨씬 효율적이었다. 또한 사람마다 표정의 다양성과 변화도가 다르기 때문에 사람에 따라 모핑에 이용될 가이드 표정을 다르게 설정하였다. 뿐만 아니라 다양한 엔진이 제작과정에 도입되는데, 이 엔진들 사이의 호환성이 완벽하지 않기 때문에 엔진간 데이터의 이동시 엔진에 따른 모델의 셋업을 재구성해야 하는 문제가 발생한다. 이러한 다양한 엔진의 사용과 엔진들간의 데이터 이동을 위한 호환성 셋업 등을 수행하는 모든 과정이 수동으로 진행된다면 3D 페이스 모델을 제작하는데 많은 비용과 시간이 소요될 수 있다. 따라서 전체 프로세스에 대한 배치 자동화와 엔진별 스크립트의 통합 등을 통해서 수동 조작되는 부분을 최소화하고 대부분의 파이프라인을 자동화해야 한다.

IV. 결 론

본 연구에서는 극사실적 표현이 적용된 페이스 모델의 제작을 위해 제작된 결과물과 비교가 용이하도록 실제 인물을 기반으로 페이스 모델을 제작하는 계획을 수립하였고 이를 위해 볼류메트릭 캡처를 통해 페이스 모델을 제작하고자 하였다. 볼류메트릭 캡처를 위해 40대의 카메라를 통해 고해상도의 이미지를 획득하였고, 획득한 이미지를 포토그래메트리를 통해 합성하는 과정을 거쳐 베이스가 되는 페이스 모델을 제작하였다. 볼류메트릭과 포토그래메트리를 통해 베이스 모델을 제작하는 경우 빠르게 초기 결과물을 생성할 수 있다는 장점이 있었으나 머리카락의 경우 합성에서 오류가 생기거나 정밀하게 구현되지 않는 문제점이 있어 후보정 과정에서 별도의 예셋을 통해 적용하는 과정이 필요하였다.

참 고 문 헌 (References)

- [1] Lee, S.-L., "The Meanings of Fashion on the Social Media of Virtual Influencer Lil Miquela," *Journal of Digital Convergence*, 19(9), pp. 323 - 333, 2021.
doi: <https://doi.org/10.14400/JDC.2021.19.9.323>
- [2] S. Hwang and M.-C. Lee, "Analysis of the Value Change of Virtual Influencers as Seen in the Press and Social Media Using Text Mining," *The Korean Journal of Advertising and Public Relations*, c23(4), pp.265-299, 2021.
- [3] N. Kumar, A. Narang and B. Lall, "Zero-Shot Normalization Driven Multi-Speaker Text to Speech Synthesis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1679-1693, 2022,
doi: <https://doi.org/10.1109/TASLP.2022.3169634>.
- [4] D. Websdale, S. Taylor and B. Milner, "Speaker-Independent Speech Animation Using Perceptual Loss Functions and Synthetic Data," *IEEE Transactions on Multimedia*, vol. 24, pp. 2539-2552, 2022,
doi: <https://doi.org/10.1109/TMM.2021.3087020>.
- [5] Darragh Higgins, Katja Zibrek, Joao Cabral, Donal Egan, Rachel McDonnell, "Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans," *Computers & Graphics*, Volume 104, pp. 116-128, ISSN 0097-8493, 2022.
doi: <https://doi.org/10.1016/j.cag.2022.03.009>.
- [6] B. Mones and S. Friedman, "Veering around the Uncanny Valley: Revealing the underlying structure of facial expressions," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 345-345, 2011.
doi: <https://doi.org/10.1109/FG.2011.5771423>.
- [7] S. Racković, C. Soares, D. Jakovetić, Z. Desnica and R. Ljubobratović, "Clustering of the Blendshape Facial Model," 2021 29th European Signal Processing Conference (EUSIPCO), pp. 1556-1560, 2021.
doi: <https://doi.org/10.23919/EUSIPCO54536.2021.9616061>.
- [8] F. Danieau et al., "Automatic Generation and Stylization of 3D Facial Rigs," 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 784-792, 2019.
doi: <https://doi.org/10.1109/VR.2019.8798208>.
- [9] Young-Ho Seo, Moonseok Oh, Gyu-Hoon Han, "The present and future of the digital human," *Broadcasting and Media Magazine*, 26(4), pp. 72-81, 2021.
- [10] M.Oh, G.-H. Han and Y.-H. Seo, "A Study on the Production Techniques of Digital Humans and Metahuman for Metaverse," *Design Research* 6, no.3, pp. 133-142, June, 2021.
doi: <https://doi.org/10.46248/kidrs.2021.3.133>
- [11] M. Oh, G.-H. Han, S.-G. Park and Y.-H. Seo. "A study on analysis of graphics pipeline for 4D volumetric capturing," *Design Research* 6, no.3, 2021 : 9-18.
doi: <https://doi.org/10.46248/kidrs.2021.3.9>
- [12] Y.-H. Seo, "Volumetric Photorealistic 4D Image Technology", *Broadcasting and Media Magazine*, 26(2), pp. 56-66, April, 2021.
- [13] B. Egger et al, "3D Morphable Face Models - Past, Present and Future," arXivLabs, Cornell University, 2020.
<https://doi.org/10.48550/arXiv.1909.01815>
- [14] F. Liu, L. Tran, X. Liu., "3D Face Modeling From Diverse Raw Scan Data," arXivLabs, Cornell University, 2019.
- [15] B. Moghaddam, J. Lee, H. Pfister and Raghu Machiraju, "Model-based 3D face capture with shape-from-silhouettes," 2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443), pp. 20-27, 2003.
doi: <https://doi.org/10.1109/AMFG.2003.1240819>

저 자 소 개



오 문 석

- 광운대학교 미디어커뮤니케이션학부 교수
- 오모션 주식회사 CEO
- 주관심분야 : 메타휴먼, 메타버스, HCI, GUI



한 규 훈

- 광운대학교 대학원 신문방송학과 석사
- 오모션 주식회사 선임연구원
- ORCID : <http://orcid.org/0000-0003-2265-3945>
- 주관심분야 : 디지털 휴먼, 메타휴먼, 메타버스, 볼류메트릭, UX/UI



서 영 호

- 1999년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2001년 2월 : 광운대학교 일반대학원 졸업(공학석사)
- 2004년 8월 : 광운대학교 일반대학원 졸업(공학박사)
- 2005년 9월 ~ 2008년 2월 : 한성대학교 조교수
- 2008년 3월 ~ 현재 : 광운대학교 전자재료공학과 교수
- ORCID : <https://orcid.org/0000-0003-1046-395X>
- 주관심분야 : 실감미디어, 2D/3D 영상 신호처리, 디지털 홀로그램