

특집논문 (Special Paper)

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

<https://doi.org/10.5909/JBE.2022.27.5.638>

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

## 메타버스를 위한 가상 휴먼의 3차원 의상 모델링

김 현 우<sup>a)</sup>, 김 동 언<sup>a)</sup>, 김 유 진<sup>a)</sup>, 박 인 규<sup>a)†</sup>

## 3D Clothes Modeling of Virtual Human for Metaverse

Hyun Woo Kim<sup>a)</sup>, Dong Eon Kim<sup>a)</sup>, Yujin Kim<sup>a)</sup>, and In Kyu Park<sup>a)†</sup>

### 요약

본 논문은 고화질 2차원 전신 영상을 입력으로 받아 영상 속 인물이 입고 있는 의상 패턴과 체형 정보를 추정한 후, 이를 반영한 3차원 가상 휴먼의 생성 기법을 제안한다. 의상의 패턴을 얻기 위해서 Cascade Mask R-CNN을 이용하여 의상 분할을 진행한다. 이후 Pix2Pix로 경계를 블러 및 배경색을 추정하고, UV-Map 기반으로 변환하여 3차원 의상 메쉬의 UV-Map을 얻을 수 있다. 또한, SMPL-X를 이용하여 체형 정보를 얻고 이를 기반으로 의상과 신체의 기본 메쉬를 변형한다. 앞서 얻은 의상 UV-Map, 체형이 반영된 의상과 신체의 메쉬를 이용해 최근 각광받고 있는 게임 엔진인 언리얼 엔진에서 렌더링하여 최종적으로 사용자가 그의 외형이 반영된 3차원 가상 휴먼의 애니메이션을 볼 수 있도록 한다.

### Abstract

In this paper, we propose the new method of creating 3D virtual-human reflecting the pattern of clothes worn by the person in the high-resolution whole body front image and the body shape data about the person. To get the pattern of clothes, we proceed Instance Segmentation and clothes parsing using Cascade Mask R-CNN. After, we use Pix2Pix to blur the boundaries and estimate the background color and can get UV-Map of 3D clothes mesh proceeding UV-Map base warping. Also, we get the body shape data using SMPL-X and deform the original clothes and body mesh. With UV-Map of clothes and deformed clothes and body mesh, user finally can see the animation of 3D virtual-human reflecting user's appearance by rendering with the state-of-the game engine, i.e. Unreal Engine.

Keyword : Metaverse, virtual human, 3D cloth modeling

a) 인하대학교 정보통신공학과(Inha University, Department of Information and Communication Engineering)

† Corresponding Author : 박인규(In Kyu Park)

E-mail: pik@inha.ac.kr

Tel: +82-32-860-9190

ORCID: <https://orcid.org/0000-0003-4774-7841>

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2022R1A4A1033549). 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2020-0-01389, 인공지능융합연구센터지원(인하대학교), No. RS-2022-00155915, 인공지능융합혁신인재양성사업(인하대학교))을 받아 수행된 연구임.

※ This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2022R1A4A1033549). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-01389, Artificial Intelligence Convergence Research Center (Inha University), No.RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)).

· Manuscript July 8, 2022; Revised August 1, 2022; Accepted August 1, 2022.

## I. 서 론

코로나 이후 온라인 활동이 증가하게 되면서 IT 업계의 수요가 매우 커졌다. 이와 함께 급부상하고 있는 것 중 하나가 바로 메타버스이다. 메타버스(Metaverse)는 가상을 의미하는 ‘Meta’와 세계를 의미하는 ‘Universe’의 합성어로, 사용자의 또 다른 자신이 존재하는 가상 세계를 의미한다.

기존 메타버스는 그래픽 수준이나 차원을 구분하지 않았지만, 그래픽 기술이 발달함에 따라 사용자들은 실제 현실과 가까운 그래픽 수준을 요구하게 되었다. 그러나 현재 메타버스 내 3차원 모델링 기법에 대한 연구가 매우 미비한 상태이다. 따라서 3차원 사물에 대한 모델링 기법과 메타버스 사용자들의 만족도를 동시에 높일 수 있는 방법에 대해 고안하였다. 그 중 대표적인 예시로 한 사용자가 메타버스 캐릭터를 생성할 때, 미리 만들어져 있는 캐릭터를 선택하는 대신 사용자의 머리와 얼굴, 의상 그리고 체형을 그대로 반영하여 메타버스 내 캐릭터로 생성할 수 있다면 3차원 모델링 기법 연구에 기여함과 동시에 사용자의 만족도를 높일 수 있을 것이다. 따라서 본 논문에서는 이 아이디어를 확장하여 사용자의 전신 영상이 입력으로 들어오면 영상 속 인물의 의상 패턴과 체형을 추정하고, 데이터를 3차원 플랫폼에서 결합한 뒤 렌더링하여 보여주는 가상 휴먼의 3차원 의상 모델링 기법을 제안한다.

본 논문의 전체적인 구성은 다음과 같다. 2장에서는 기존의 의상 복원 기술에 대해 소개하고, 3장에서는 본 논문에서 단계별로 이용하기 위해 선정한 개발 환경과 프로그램을 소개한다. 4장에서는 제안하는 가상 휴먼의 3차원 의상 모델링 기법을 자세히 기술한다. 5장에서는 구현한 기법의 데모 실행 결과를 보여주고, 결과에 대한 고찰을 기술한다. 마지막으로 6장에서 논문의 결론을 맺는다.

## II. 관련 연구

2차원 영상으로부터 3차원 메쉬를 생성하는 연구는 딥러닝 기술의 발달 덕분에 성능과 효율성이 크게 개선되어 활발히 진행되고 있다. 해당 기술에 대한 연구는 크게 두 가지로 나뉘는데, 첫 번째는 기본 메쉬 없이 2차원 영상만으로

새로운 메쉬를 직접 생성하는 방법이다. 하지만 이 방법은 3차원 데이터셋의 부족으로 실용성이 떨어져 아직 대중적으로 사용되지는 못하고 있다. 두 번째는 기본 메쉬를 기반으로 하여 파라미터를 추정한 뒤 기본 메쉬를 변형하는 방법이다. 대표적으로 인물 사진으로부터 3차원 신체로 복원하는 모델인 SMPL이 알려져 있다. SMPL (Skinned Multi-Person Linear Model)은 신체의 체형과 자세를 동시에 정확하게 추정하는 3차원 학습 모델로, 합성곱 신경망으로 2차원 관절을 추정한 뒤 자세 파라미터와 체형 파라미터를 추정하여 기본 체형 메쉬를 변형하는 방법을 이용하고 있다<sup>[1]</sup>.

2차원 영상으로부터 3차원 메쉬의 UV-Map을 생성하는 연구는 딥러닝 기술의 발전으로 다양한 분야에 적용이 가능해졌는데, 특히 영상 속 인물의 얼굴 UV-Map을 추정하는 연구들이 활발하게 이루어지고 있다. UV-GAN은 생성적 적대 신경망을 이용하여 입력 영상 속 인물의 얼굴 UV-Map을 생성하는 기법을 제안하였다<sup>[2]</sup>. 이는 다른 분야에서도 3차원 메쉬의 UV-Map을 딥러닝 기술을 이용하여 충분히 복원 가능하다는 것을 알 수 있다. 하지만 본 논문에서 의상에 대해서 UV-Map 생성 시 생성적 적대 신경망만을 이용하기에는 얼굴에 비해 양질의 데이터셋이 부족하며, 의상의 패턴 특징이 다양하여 추정에 어려움이 있다. 그러므로 본 논문에서는 3차원 메쉬의 UV-Map을 추정하는 것은 동일하나 생성적 적대 신경망만을 이용하는 것이 아닌 크기 조정, 분할 등 다양한 기법을 함께 이용하여 UV-Map을 생성한다.

본 논문의 목적과 가장 유사한 연구로는 VITON (Virtual Try-On) 기법이 있다<sup>[3]</sup>. VITON은 인물과 의상의 2차원 영상을 입력으로 주었을 때, 따로 주어진 의상을 타겟으로 하여 인물 사진의 의상을 대체하여 씌워진 인물 영상을 출력하는 방식을 이용한다. 의상을 변환하기 위해 인물의 자세, 체형, 그리고 얼굴에 대한 정보를 추출하여 여러 기능에 대한 정보가 모두 담긴 하나의 인물 표현 데이터를 생성한다. 인물 표현 데이터를 기반으로, TPS 변환을 이용하여 타겟 의상을 변환하고 최종적으로 인물이 입고 있는 의상의 마스크와 일치하는 타겟 의상 이미지를 생성한다. VITON은 의상 복원 연구에서 뛰어난 성능을 보이지만, 2차원 인물의 의상을 대체하는 것이므로 메타버스 같은 3차원 플랫폼에

서 활용할 수 없다. 움직이는 가상 휴먼을 위한 의상 복원 기법을 연구하기 위해서는 다량의 2차원 영상 혹은 하나의 동영상을 입력으로 넣어 일정 시간 간격으로 VITON 네트워크를 실시간으로 수행하여 화면에 보여주는 방식을 이용하여 개선할 수 있다.

본 논문에서는 기존의 의상 3차원 복원 연구에서 나아가 3차원 플랫폼 상에서 움직이는 가상 휴먼을 생성하기 위해 체형 추정 모델, 의상 패턴 추정 모델 등 여러 모델을 조합한다. 그리고 3차원 편집 프로그램에서 메쉬를 변형하고, 3차원 오픈소스 플랫폼에서 제공하는 기본 에셋을 활용하여 최종적으로 3차원으로 렌더링된 가상 휴먼을 생성한다.

### III. 개발 환경 및 도구

본 논문에 들어가기에 앞서, 의상과 신체 변형에 이용할 기본 모델과 변형 및 3차원 렌더링을 진행하기에 적합한 개발 환경에 대한 조사를 수행하였다. 여러 후보 프로그램을 고려한 결과 그림 1과 같이 결정하였다.

우선 의상과 신체의 기본 모델은 Epic Games에서 제작한 3차원 가상 휴먼 프레임워크인 메타휴먼 크리에이터(MetaHuman Creator)에서 제공하는 파일을 이용하였다<sup>[4]</sup>.

메타휴먼 크리에이터에서는 기본적인 의상과 평균 체형에 대한 메쉬와 스켈레톤을 무료로 제공하고 있다. 이 메쉬와 스켈레톤은 이후에 기술할 3차원 가상인물 렌더링 플랫폼과 같은 형식을 갖추어야 하는데, 마찬가지로 Epic Games에서 제작한 3차원 플랫폼인 언리얼 엔진(Unreal Engine)을 이용하였다<sup>[5]</sup>. 메타휴먼 크리에이터와 언리얼 엔진은 연구 목적에 한해 무료로 이용 가능하며 같은 파일 포맷을 공유하기 때문에 본 논문에서 활용하기에 매우 적합하다고 판단하였다. 그리고 메쉬를 변형할 개발 환경으로는 오픈 소스 3차원 컴퓨터 그래픽 제작 소프트웨어인 블렌더를 선택하였다<sup>[6]</sup>. 블렌더는 정점으로 구성된 3차원 객체를 쉽게 편집할 수 있고, 물체 변형을 위한 다양한 내장 함수를 제공하고 있다. 또한, 복잡하거나 정점이 많은 객체에 대하여 빠르게 연산할 수 있도록 파이썬 API를 제공하고 있다.

### IV. 제안하는 가상 휴먼의 3차원 의상 모델링 기법

본 장에서는 사용자의 정면 모습이 담긴 고해상도 2차원 영상이 입력되었을 때 사용자의 의상 패턴, 체형, 얼굴을 추정하여 3차원 플랫폼에서 가상 휴먼으로 렌더링하는 3차



그림 1. 메타휴먼 크리에이터(좌)와 블렌더(우)  
Fig. 1. MetaHuman Creator(left) and Blender(right)

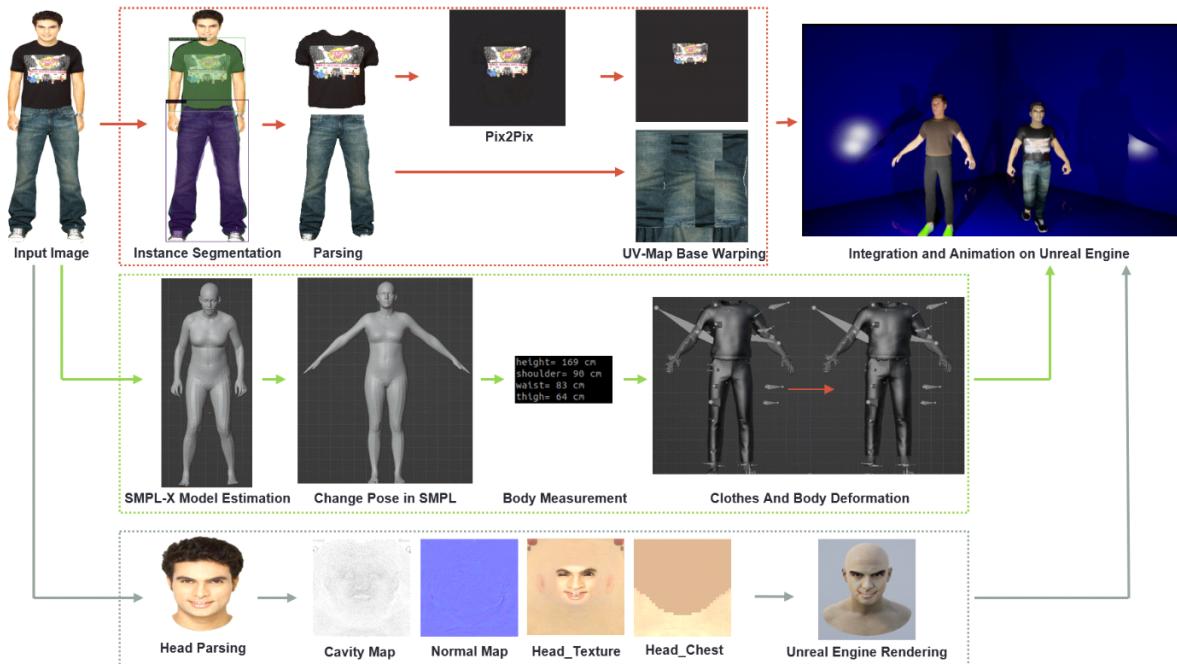


그림 2. 제안하는 3차원 의상 모델링 및 체형 추정의 전체 파이프라인  
 Fig. 2. Overview of the proposed system

원 의상 모델링 기법에 대해 자세히 서술한다. 제안하는 기법은 체형 추정, 의상 패턴 추정, 3차원 렌더링 및 애니메이션의 세 단계로 나눠진다. 제안하는 3차원 의상 모델링 기법의 구조는 그림 2에 도시하였다. 얼굴 부분에 대해서 인하대학교 시각컴퓨팅 및 학습 연구실의 지원을 받아, 진행 할 데모에 대한 얼굴 머티리얼을 사전에 제공받아 프로젝트에 이용하였다.

### 1. SMPL 기반 체형 추정과 메쉬 변형

첫 번째 단계에 해당하는 입력 영상 속 인물의 체형 정보를 추정하는 과정을 설명한다. 단일 영상으로부터 인물의 신체를 3차원 모델로 복원하는 방법은 매우 어렵고 까다로운데, 이러한 문제를 사전에 정의된 체형, 자세 정보들을 이용하여 복원하는 SMPL을 이용하여 해결하였다. SMPL (Skinned Multi-Person Linear Model)은 신체의 체형과 자세를 동시에 정확하게 추정하는 3차원 학습 모델이다. 사전에 정의된 체형과 자세 정보를 통해 입력 영상 속 인물로부터 체형 파라미터(shape parameter)와 자세 파라미터(pose

parameter) 값을 추정한 뒤 3차원 신체 모델을 생성한다.

나아가 SMPL-X (Skinned Multi-Person Linear Model eXpressive)는 인물의 표정과 손, 발의 자세한 자세를 추정하는 SMPL의 확장 모델이다. 본 논문에서는 자세한 체형 정보를 얻어야 하기 때문에 SMPL-X가 더 좋은 성능을 보일 것이라 판단하여 최초 신체 모델을 생성하는 과정에서 이를 이용하였다<sup>[7]</sup>.

그림 3과 같이 입력 영상 속 인물의 자세와 체형을 추정한 뒤에, 신체 주요 부위 측정을 위해 자세를 조정해야 한다. 이 과정은 SMPL이 좀 더 활용하기 용이하므로 SMPL-X 모델에서 SMPL 모델로의 변환을 수행한다. 여기서 체형 파라미터는 그대로 이동하는 반면, 신체 측정을 위해서 곧은 자세로 변환하기 위해 특정 자세 파라미터의 값을 바꾸어 주었다. 이 작업을 수행하면 체형 파라미터는 변화되지 않고 자세만 수정된 SMPL 모델을 얻을 수 있다.

마지막으로 자세를 수정한 SMPL 모델을 이용해 신체의 주요 부위에 대한 치수를 측정한다. SMPL 모델은 정점들 (vertices)의 집합으로 이루어진 모델이다. 이러한 특징에서 착안하여, 특정 면(face)과 점(vertex) 사이의 거리를 계산하

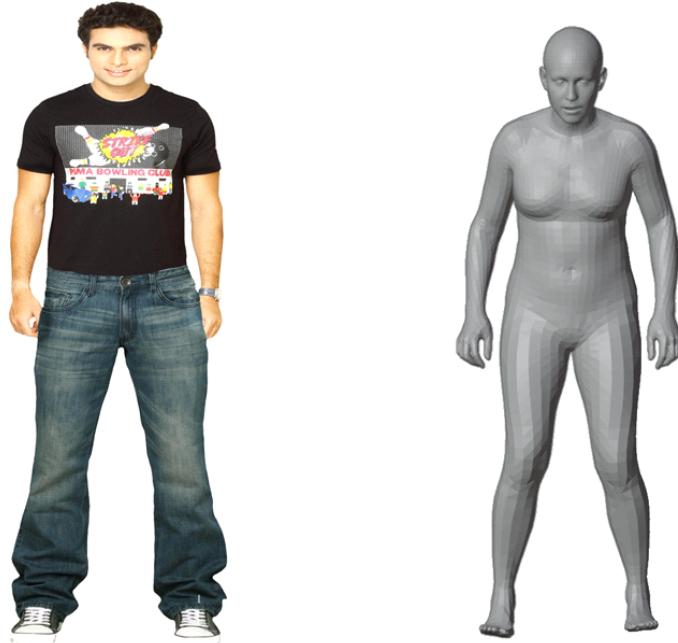


그림 3. SMPL-X를 활용한 체형과 자세 추정  
Fig. 3. Estimation of shape and pose using SMPL-X

여 특정 부위의 치수를 측정하는 알고리즘을 활용하였다. 파이썬 언어의 PyWavefront 패키지를 이용하여 obj 파일을 면과 점으로 구성된 포맷으로 변환한 뒤, Trimesh 패키지를 이용하여 주요 신체 부위의 치수를 측정한다. 측정할 신체 주요 부위로는 키, 어깨, 허리, 허벅지를 선택했는데, 키의

경우 바운딩 박스의 Y 좌표의 최댓값과 최솟값의 차이이고, 어깨나 허리, 허벅지의 경우 전체 키에서 해당 부위가 위치하는 Y좌표 값을 지정한 뒤 해당 좌표의 단면 둘레를 계산하여 값을 얻을 수 있다.

체형 추정을 통해 얻은 신체 주요 부위의 측정값을 텍스

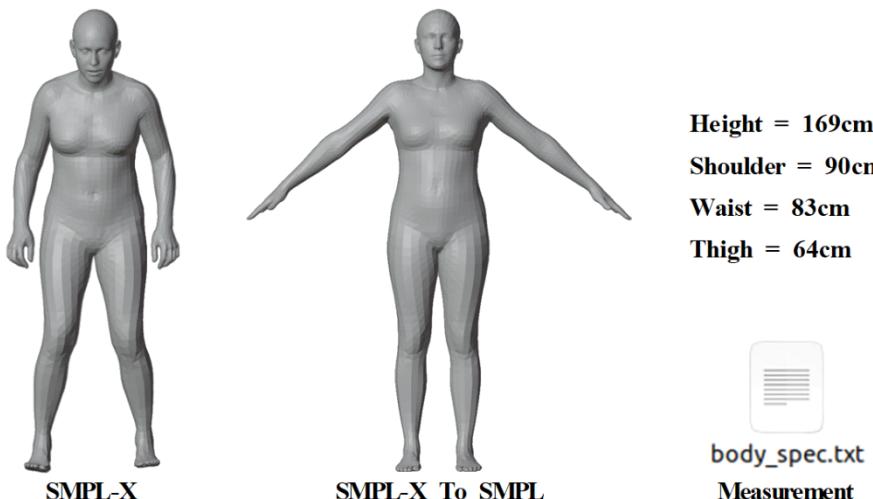


그림 4. SMPL을 이용한 신체 주요부위 치수 측정  
Fig. 4. Body measurement using SMPL

트 파일로 저장하여 최종적으로 이 값을 이후 의상과 신체 메쉬를 변형하는 데 이용할 수 있도록 하였다. 체형 추정의 전체적인 파이프라인은 그림 4에 도시하였다.

SMPL 모델에서 얻은 체형 추정 정보를 이용하여 기본적인 신체, 의상 메쉬를 변형시킨다. 메타휴먼 크리에이터에서 제공하는 신체, 의상 파일은 메쉬와 스켈레톤으로 구성되어 있고, 각 스켈레톤마다 영향을 받는 정점 집합이 정의되어 있다. 예를 들어 허리에 해당하는 스켈레톤들은 해당 스켈레톤이 회전할 때 함께 움직이는 메쉬의 정점 집합이 사전에 정의된다. 이 특징을 이용하여, 각 신체 부위에 해당하는 스켈레톤 집합을 미리 정의한 뒤 스켈레톤 집합에 정의되어 있는 모든 정점 집합을 계산하여 선택한다. 그림 5는 각 신체 부위에 해당하는 모든 정점 집합을 시각적으로 표시하여 도시한 것이다.

그 후, 선택된 정점 집합에 대해 변형을 수행한다. 블렌더에서 제공하는 내장 함수 중에서 키를 조정하는 작업은 원점을 기준으로 Z축 방향으로 스케일링하는 함수를 이용하였고, 키 이외의 신체 부위에 대해서는 법선 벡터를 기준으로 면을 축소하거나 살찌우는 `shrink_fatten` 함수를 이용하였다.

블렌더 파이썬 스크립트를 수행할 기본 모델로는 평균

체형의 남성 모델을 이용하였고, 이 모델에 대한 신체 부위 치수 값은 메타휴먼 크리에이터에서 제공하고 있지 않기 때문에 여러 입력 영상에 대해 직접 변환을 수행해본 뒤, 가장 적합한 값으로 선정하였다. 평균 체형 남성의 신체 주요 부위 측정값과 앞서 체형 추정을 통해 얻은 측정값을 이용하여 평균 체형에 대한 차이의 비율을 계산하고, 평균 체형 모델의 메쉬를 스케일링하거나 내장 함수를 이용하여 축소 및 살찌우는 변형 작업을 진행한다. 이 스크립트를 평균 남성의 의상과 신체 메쉬에 대해 수행하면, 그림 6과 같이 입력 영상 속 인물의 체형 정보가 반영된 의상과 신체 메쉬를 얻을 수 있게 된다.

## 2. 의상 분할 및 패턴 추정

두 번째 단계인 의상의 UV-Map을 생성하는 과정은 의상 분할로부터 시작된다. 의상 분할을 위해 MMDetection<sup>[8]</sup>으로 구현된 Instance Segmentation 모델인 Cascade Mask R-CNN을 이용한다<sup>[9]</sup>. 학습에 필요한 데이터셋은 DeepFashion2<sup>[10]</sup>를 이용한다. DeepFashion2는 13개의 의상 카테고리가 있으며, 학습 데이터셋은 191,961장, 검증 데이터셋은 32,153장을 제공한다. 이때 데이터셋을 이용하여 효

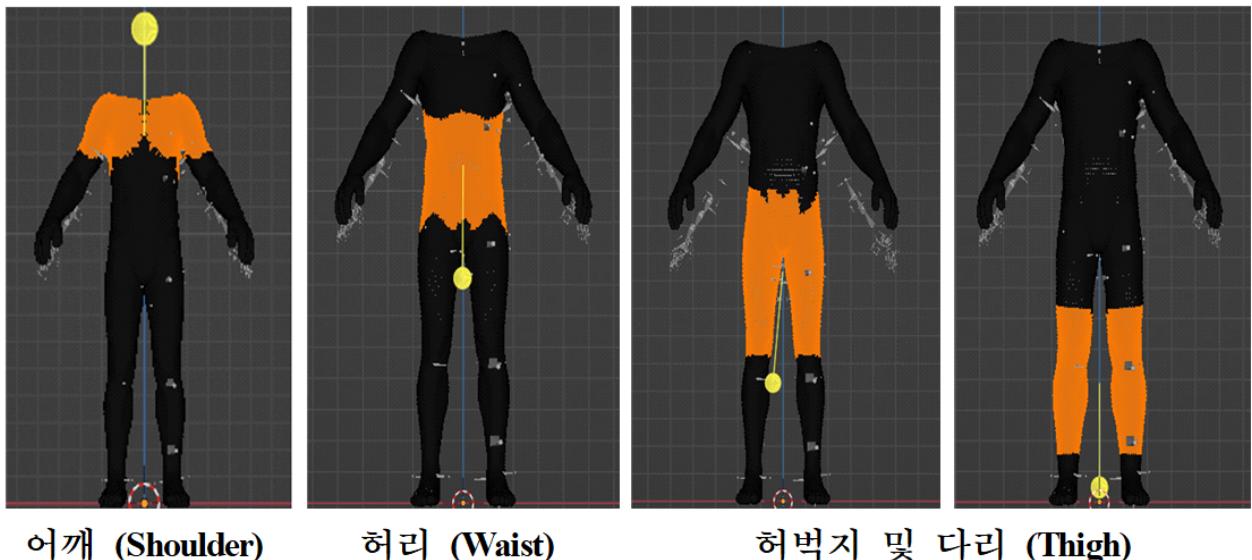


그림 5. 각 신체 부위별 휴먼 메쉬의 정점 그룹 선택 결과  
Fig. 5. Selection of human mesh vertex group

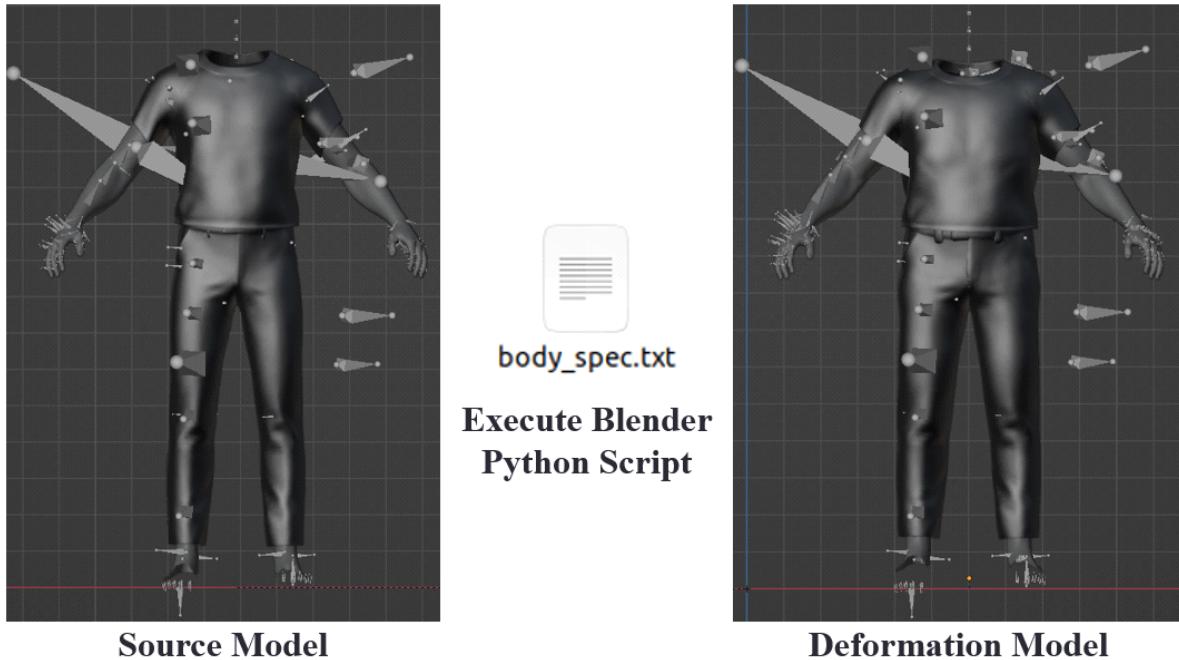


그림 6. 의상과 체형 메쉬 변형  
Fig. 6. Clothes and body mesh deformation

율적인 학습을 진행하기 위해 이미지의 해상도를  $500 \times 500$ 로 크기 조정과 정규화를 진행한 후, 수평과 수직 방향 플립을 0.5의 확률로 수행하는 데이터 증강을 진행한다. 학습을 위해 batch size를 8로 설정하였다. 옵티마이저는 Adam<sup>[11]</sup>을 이용하여 학습률은 0.000002로 설정하고 epoch 6번을 진행했다.

의상 분할의 실행결과를 원하는 목적에 맞게 이용하기 위해 마스크를 적용하여 관심 영역을 제외한 나머지 부분은 모두 흰색으로 화소 값을 변경하였으며, 그 후에 취득된 경계를 이용하여 의상의 영역을 취득하였다. 위 모델을 이용하여 그림 7과 같이 의상의 경계를 분할하였으며 배경이 있는 경우에도 분할은 정상적으로 진행이 되었지만, 배경 색이 분할 경계에 남아있는 것을 알 수 있다. 위 모델의 출력결과를 이용하여 의상의 영역만 추출한 후에, 경계에 대한 블러와 배경색을 추정하기 위해 Image-to-Image translation 모델인 Pix2Pix를 이용한다. Pix2Pix는 Image-to-Image translation만을 수행하는 GAN 모델로, 입력 이미지에서의 경계와 배경색을 추정하여 입력 이미지의 특성을

담고 있는 새로운 타겟 이미지를 만들기 적합한 모델이다<sup>[12]</sup>.

하의 부분에 대해서는 구축할 수 있는 데이터셋이 충분하지 않아 상의 영역에 대해서만 Pix2Pix를 이용한다. 학습에 필요한 데이터셋 확보를 위해 Kaggle에서 고해상도 의상 이미지를 제공하는 Fashion Product Images Dataset을 이용하였다<sup>[13]</sup>. 기존 고해상도 이미지를  $1024 \times 1024$ 으로 크기를 변경하였고, 의상의 패턴을 제외한 나머지 부분에 대해서는 지배적인 색상을 가지도록 데이터셋을 구축하였다. 경계에 대해서도 블러 처리를 통해서 부드러운 경계면을 가질 수 있도록 했다. 그림 8과 같이 긴 상의, 짧은 상의에 대해서 데이터셋을 구축하였다. 이때 긴 상의는 총 75장이며, 짧은 상의는 총 175장이다.

이후 모델의 학습을 진행한 뒤에 실제 입력 이미지에 대해서 추론을 위해 분할 모델로 추출한 의상을  $1024 \times 1024$  흰색 이미지 가운데에 합성을 진행하였다. 이때 분할된 의상 이미지의 사이즈를  $650 \times 650$ 으로 이미지 크기를 변경하였다.



그림 7. Cascade Mask R-CNN을 이용한 의상 분할  
Fig. 7. Clothes segmentation using Cascade Mask R-CNN



그림 8. Pix2Pix로 생성된 데이터셋  
Fig. 8. Dataset generated by Pix2Pix



그림 9. 입력 영상 속 의상 분할 파이프라인  
Fig. 9. Overview of clothes parsing of input image

그림 9는 입력 영상에 대해서 의상을 분할한 뒤에 Pix2Pix를 통과시킨 결과이다. 의상의 경계 블러와 배경색에 대한 추정이 정상적으로 이루어지는 것을 볼 수 있다. 사용하는 의상은 3차원 객체이므로 블렌더에서 UV-Map의 형식을 추출할 수 있다. UV-Map이란 텍스쳐를 쉽게 감싸는데 사용되는 3차원 모델의 평면으로, 이를 이용하면 의상의 패턴을 추출한 것을 3차원 모델의 표면에 생성할 수 있다.

의상의 UV-Map은 그림 10과 같으며, 상의의 경우 Pix2Pix로 생성된 이미지를 이용하며, 하의의 경우 의상 분할을 통해 생성된 이미지를 이용한다. 상의의 경우 긴 상의를 기반으로 하여 짧은 상의가 생성된 것으로 유사한 UV-Map을 가지고 있는 것을 알 수 있다. 그러므로 Pix2Pix의 출력 결과를 이용하여 가운데 패턴 추출 후 이미지 크기 조정을 수행한다. 이후 기존 출력 결과의 배경색에 대해 부드러운 색 표현을 위하여 정규화를 진행한 뒤에 UV-Map의 상의 정면 부분에 합성을 수행한다. 하의의 경우 상의보다 복잡한 변환을 수행한다. 하의는 사타구니 지점을 찾는 것부터

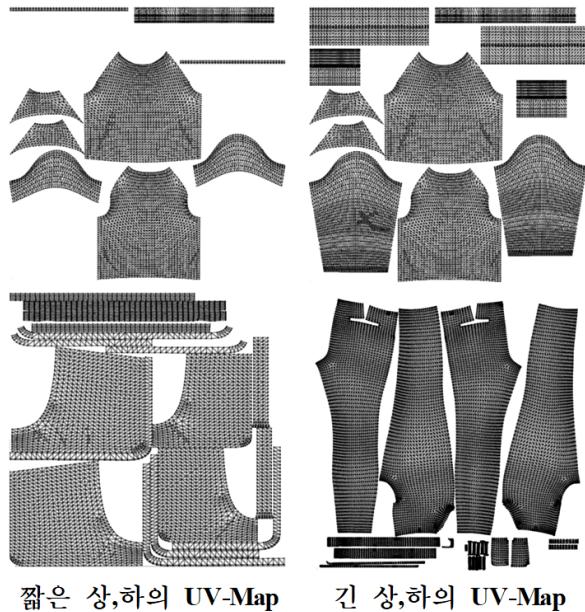


그림 10. 상의와 하의의 UV-Map  
Fig. 10. UV-Map of top and bottom



그림 11. UV-Map 기반 상의와 하의 왜곡 결과  
Fig. 11. Results of warping UV-Map of top and bottom

시작되는데, 분할된 의상을 그대로 이용하므로 의상의 내부영역이 아닌 경우 화소 값이 일정하다는 것과 하의의 특성을 이용하여 가운데부터 탐색을 시작해 사타구니 지점을 구하게 된다. 이후 사타구니를 기준으로 하의 이미지를 총 4개의 영역으로 분할을 진행한 후, 분할된 영역에서 4개의 꼭짓점을 특정한 뒤에 UV-Map에 해당하는 부분으로 변환을 진행한다. 반팔과 긴 하의의 변환을 진행한 결과는 그림 11에 도시하였다.

### 3. 3차원 렌더링과 애니메이션

마지막으로 의상, 체형, 얼굴에 대한 추정 데이터들을 이용하여 3차원으로 렌더링된 가상 휴먼을 보여준다. 이 과정은 앞에서 언급한 언리얼 엔진에서 진행한다. 언리얼 엔진은 혼존하는 3차원 게임 엔진 중에서 높은 퍼포먼스로 가장 많이 각광받고 있다. 또한, 기본적으로 제공하는 여러 시각 효과들을 부여하여 실제 가상세계 속에서 만든 3차원 가상 휴먼의 움직이는 모습을 확인할 수 있기에 본 논문의 최종 결과물인 3차원으로 렌더링된 가상 휴먼의 애니메이션을 구현하기에 가장 적합한 플랫폼이라고 판단하였다.

앞서 얻은 의상 추정 UV-Map과 체형 추정이 반영된 의

상, 신체 메쉬를 언리얼 엔진으로 전송한 뒤, 3차원으로 렌더링한 후 언리얼 엔진에서 제공하는 무료 애니메이션을 적용하여 3차원으로 렌더링된 가상 휴먼의 애니메이션을 구현할 수 있다. 영상물의 퀄리티를 높이기 위해 그림 12와 같이 전역 조명, 그림자, 안티 에일리어싱 등의 시각 효과를 추가하였다.

## V. 실험 결과

본 장에서는 위에서 기술한 프로그램의 성능을 테스트하기 위해 데모 이미지를 이용하여 각 단계별 수행 결과를 보여주고 분석하여 문제점과 개선점을 기술한다.

### 1. 데모 실행 결과

입력 영상의 크기는  $3000 \times 3000$ 의 해상도를 가지고 있으며, 총 2개의 데모 이미지를 이용하였다. 각각 순서대로 덩치 있고 키가 작으며 반팔과 긴 바지를 착용한 남성, 마르고 키가 크며 반팔과 반바지를 착용한 남성을 선정하였다. 이에 대한 실행 결과는 그림 13에 도시하였다.

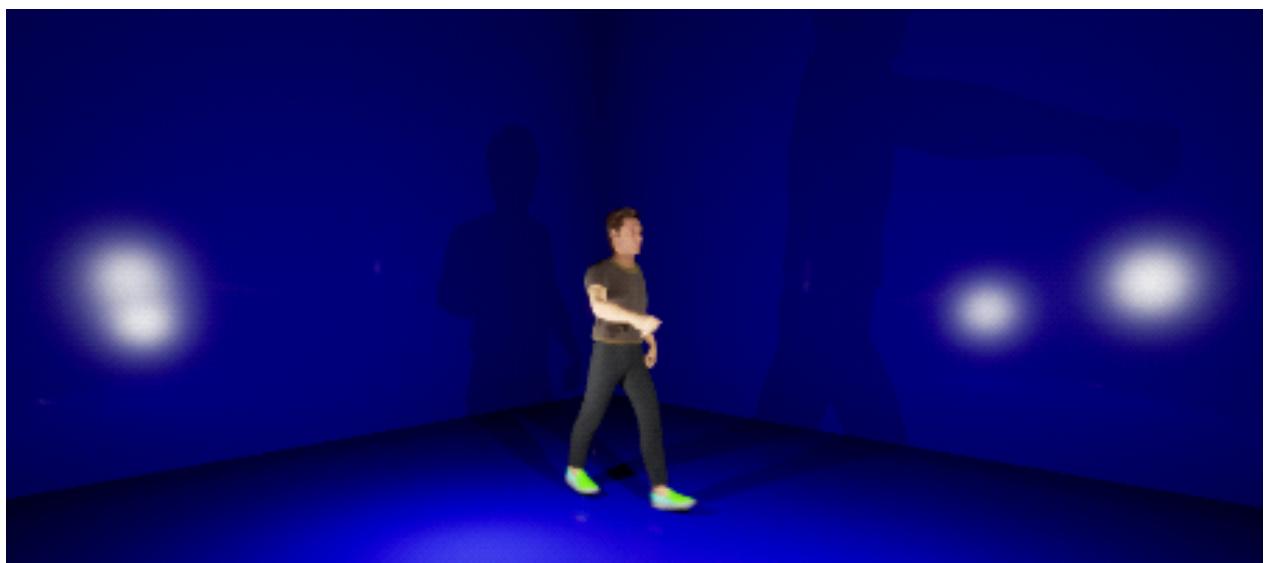


그림 12. 3차원 렌더링과 애니메이션 예시

Fig. 12. An example of 3D rendering and animation

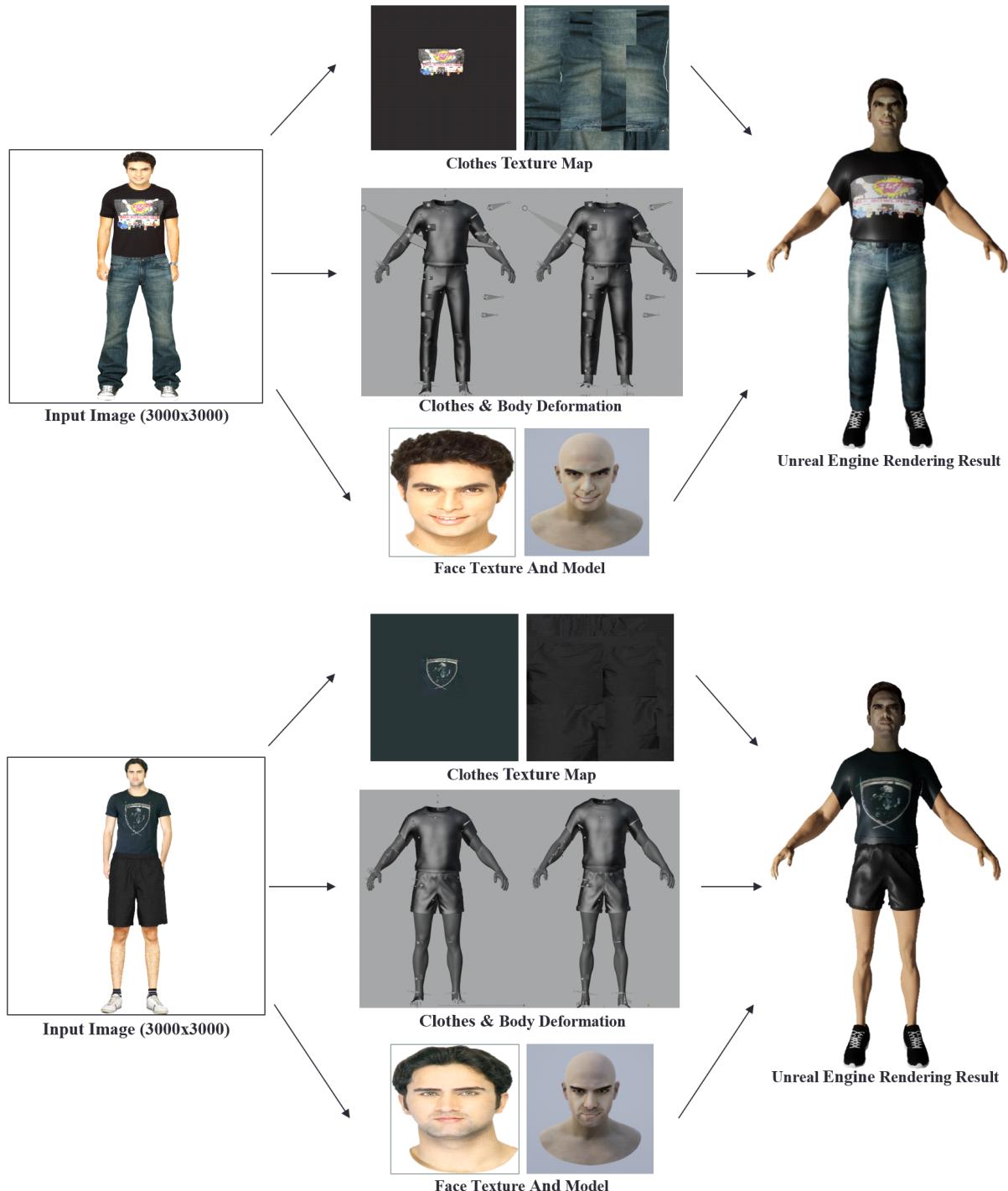


그림 13. 입력 영상에 따른 실행 결과  
Fig. 13. Results according to input images

## 2. 프로그램의 한계와 고찰

데모 이미지에 대한 결과를 확인하면 입력 이미지 의상 패턴과 체형이 언리얼 엔진 상의 3차원 가상 휴먼에게 반영이 된 것을 볼 수 있다. 하지만 의상 메쉬를 변형하는 과정에서 과도하게 변형될 경우 메쉬가 부자연스럽게 변형되는 현상을 확인할 수 있었다. 그림 13의 두 번째 사진에 대한 실행 결과에서, 짧은 하의 메쉬의 모서리 부분이 부드럽지 않게 변형된 것을 확인하였다. 이는 단면으로 구성된 메쉬를 가정하고 블렌더의 내장 함수로 변형하도록 코드를 구현했는데, 메타휴먼 크리에이터에서 하의의 가장 끝이 비

교적 두꺼운 것을 3차원에서 표현하기 위해 여러 면으로 만들어서 발생한 현상이다. 의상 메쉬에서 필요 이상으로 구성된 면을 최대한 제거하면 해결이 가능하지만, 의상 메쉬의 품질이 낮아지기 때문에 그 사이의 적절한 지점을 결정하여 이용하면 해결될 것으로 예상된다.

체형을 추정하는 과정에서도 특정 입력 영상에 대해 체형을 올바르게 추정하지 못하는 예시들이 존재했다. SMPL-X 역시 체형을 추정하는 데 좋은 성능을 보여주는 모델이지만 본 프로그램 목표에서 요구하는 자세한 체형 정보를 얻기에는 명확한 한계가 분명히 존재했고, 키가 크거나 덩치가 큰 남성처럼 데이터가 부족한 체형의 경우 제대로 예측

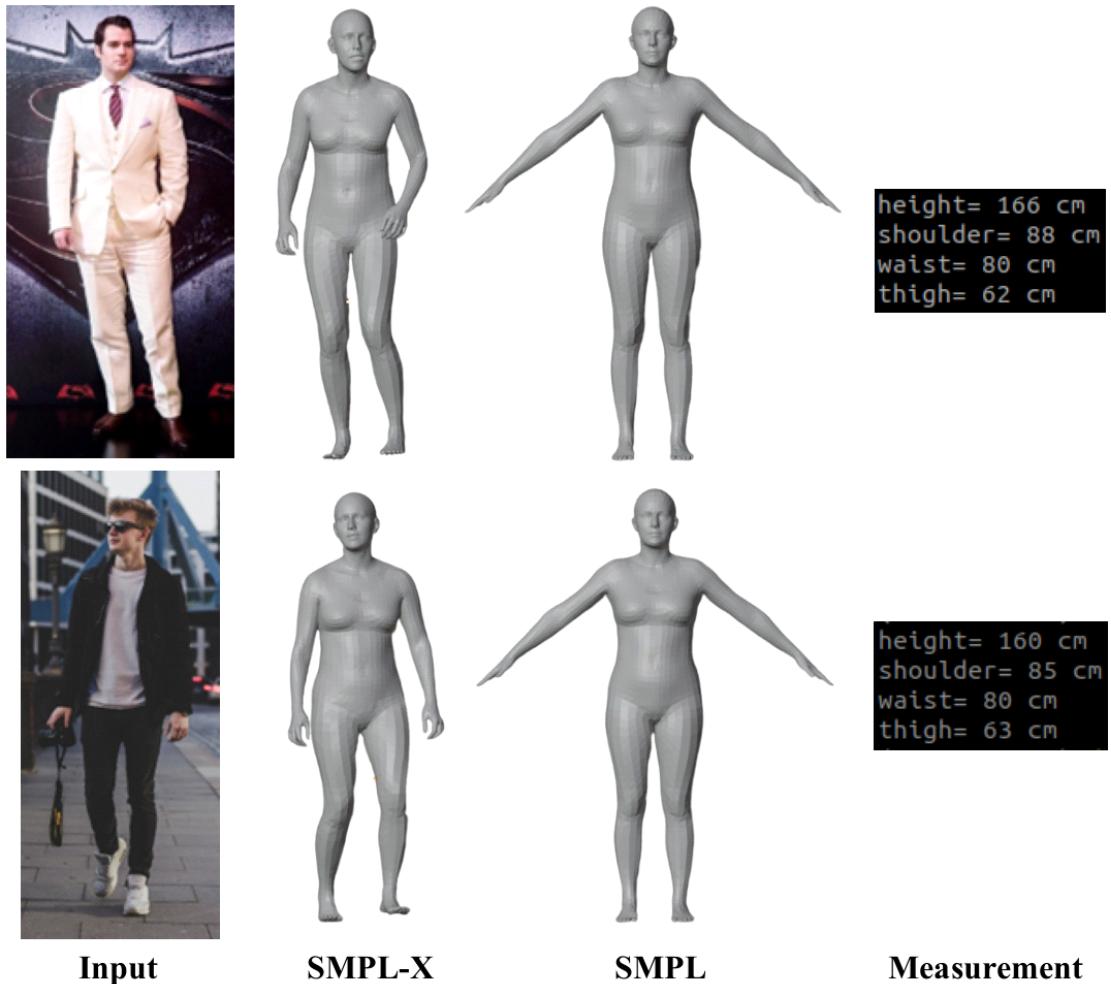


그림 14. 체형 추정 단계에서 실패한 예시 결과

Fig. 14. Failure cases on estimating body shape



그림 15. 의상 분할 및 패턴 추정 단계에서 실패한 예시 결과  
 Fig. 15. Failure cases on clothes segmentation and pattern estimation

하지 못하는 것을 확인할 수 있었다. 그에 대한 예시로, 그림 14의 첫 번째 사진의 경우 키가 평균 이상이고 덩치가 큰 남성이지만 두 번째 사진에 비해 키, 허리와 어깨에 대한 치수 측정 결과가 나쁘게 측정된 것을 확인할 수 있었다. 이는 SMPL-X 모델의 학습 데이터를 늘릴 수 있다면 해결이 가능할 것이다.

마지막으로 의상의 UV-Map을 생성하는 과정에서도 아쉬운 성능을 보여주었다. UV-Map을 생성하는 과정 중 Image-to-Image translation, UV-Map 기반 변환, 이미지 사이즈 변경을 수행하기 때문에 원본 화소 값들이 훼손되는 문제가 있어 블러 효과가 생긴 것을 알 수 있다. 그리고 Image-to-Image translation 과정에 의상 패턴이 복잡해지면 제대로 된 결과를 도출하지 못하는 경우를 확인하였다. 이는 데이터셋 확보 및 진보된 GAN 모델을 이용한다면 개선 가능하다. UV-Map 기반 변환의 경우에도 특징점을 현재처럼 naïve하게 찾는 방식을 이용하게 된다면 사타구니 지점을 찾지 못하면 오류가 발생하게 된다. 또한 그림 15와 같이 의상분할 과정에서 카테고리 추정과 분할을 제대로 진행하지 못하면 짧은 하의를 입는 인물의 사진을 긴 하의를 입은 인물로 판단하여 UV-Map이 잘못 만들어지는 것을 알 수 있다. 상의의 경우에도 색의 정규화가 제대로 진행되지 않으면 경계에서 부자연스러운 느낌을 받을 수 있고, 그림 15의 오른쪽 인물의 상의와 같이 큰 디테일을 가져올 수 있지만, 섬세한 디테일은 UV-Map에 표현하지 못하는 한계가 있다.

## VI. 결 론

본 논문에서는 메타버스 사용자들을 대상으로 하여 입력 영상 속 인물의 의상, 체형, 얼굴에 대한 정보를 취득하는 모델을 이용하여 3차원 플랫폼 내에서 움직이는 가상 휴먼을 생성하는 방법을 제안하고 그에 대한 성능을 확인한다. 프로그램의 결과물을 언리얼 엔진을 이용하여 실제 메타버스 내에서 사용자의 외형을 그대로 반영한 3차원 가상 캐릭터를 생성하고, 실제 사람처럼 움직이는 모습을 확인할 수 있었다. 블렌더나 언리얼 엔진 이외의 3 차원 객체 편집 프로그램을 활용하여 다양한 의상의 카테고리를 제작하고, 본 프로그램을 경량화할 수 있을 것이다. 본 논문에서 설계한 프로그램은 메타버스를 이용하는 모든 사용자들의 만족도를 향상시킬 뿐만 아니라, 실제 가상현실이나 메타버스 내에서 활용되는 3 차원 모델링 기법 연구에 도움이 될 것이라 예상한다.

## 참 고 문 헌 (References)

- [1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A Skinned multi-person linear model,” ACM Trans. on Graphics, vol. 34, no. 6, pp. 1-16, 2015.  
 doi: <https://doi.org/10.1145/2816795.2818013>
- [2] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, “UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition,” arXiv preprint arXiv:1712.04695, 2017.  
 doi: <https://doi.org/10.48550/arXiv.1712.04695>

- [3] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "VITON: An image-based virtual try-on network," arXiv preprint arXiv:1711.08447, 2017.  
doi: <https://doi.org/10.48550/arXiv.1711.08447>
- [4] Metahuman Creator, <https://www.unrealengine.com/ko/metahuman> (accessed June 13, 2022)
- [5] Unreal Engine, <https://www.unrealengine.com/> (accessed June 13, 2022)
- [6] Blender, <https://www.blender.org> (accessed June 13, 2022)
- [7] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas, and M. J. Black, "Expressive body capture: 3D hands, face, and body from a single image," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2019.  
doi: <https://doi.org/10.1109/CVPR.2019.01123>
- [8] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "MMDetection: Open MMLab detection toolbox and benchmark," arXiv preprint arXiv:1906.07155, 2019.  
doi: <https://doi.org/10.48550/arXiv.1906.07155>
- [9] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High quality object detection and instance segmentation," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 43, no. 5, pp. 1483-1498, 2019.  
doi: <https://doi.org/10.1109/TPAMI.2019.2956516>
- [10] Y. Ge, R. Zhang, X. Wang, X. Tang, and P. Luo, "DeepFashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2019.  
doi: <https://doi.org/10.1109/CVPR.2019.00548>
- [11] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.  
doi: <https://doi.org/10.48550/arXiv.1412.6980>
- [12] P. Isola, J. -Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," Proc. IEEE Conference on Computer Vision and Pattern Recognition, June 2017.  
doi: <https://doi.org/10.1109/CVPR.2017.632>
- [13] Fashion Product Images Dataset, <https://www.kaggle.com/datasets/paramagarwal/fashion-product-images-dataset> (accessed June 13, 2022)

---

## 저자 소개

---

### 김현우



- 2017년 2월 ~ 현재 : 인하대학교 정보통신공학과 학사과정  
- ORCID : <https://orcid.org/0000-0002-9796-0029>  
- 주관심분야 : 컴퓨터비전 및 그래픽스, deep learning

### 김동언



- 2017년 2월 ~ 현재 : 인하대학교 정보통신공학과 학사과정  
- ORCID : <https://orcid.org/0000-0003-4731-3461>  
- 주관심분야 : 알고리즘, deep learning

---

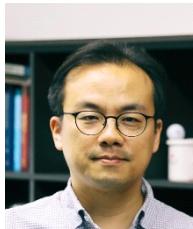
저자소개

---



김 유 진

- 2020년 8월 : 인하대학교 정보통신공학과 학사
- 2022년 8월 : 인하대학교 전기컴퓨터공학과 석사
- ORCID : <https://orcid.org/0000-0002-1990-2704>
- 주관심분야 : 컴퓨터비전 및 그래픽스, deep learning, face synthesis



박 인 규

- 1995년 2월 : 서울대학교 제어계측공학과 학사
- 1997년 2월 : 서울대학교 제어계측공학과 석사
- 2001년 8월 : 서울대학교 전기컴퓨터공학부 박사
- 2001년 9월 ~ 2004년 2월 : 삼성중합기술원 멀티미디어랩 전문연구원
- 2007년 1월 ~ 2008년 2월 : Mitsubishi Electric Research Laboratories (MERL) 방문연구원
- 2014년 9월 ~ 2015년 8월 : MIT Media Lab 방문부교수
- 2018년 7월 ~ 2019년 6월 : University of California, San Diego (UCSD) 방문학자
- 2004년 3월 ~ 현재 : 인하대학교 정보통신공학과 교수
- 2020년 4월 ~ 현재 : 인하대학교 인공지능융합연구센터 센터장
- ORCID : <https://orcid.org/0000-0003-4774-7841>
- 주관심분야 : 컴퓨터비전 및 그래픽스, deep learning, GPGPU