

## 딥러닝 기반 자세 추출을 통한 메타버스 가상 피팅 기술 구현

이범로<sup>o</sup>, 이상원\*, 신수진\*

<sup>o</sup>청강문화산업대학교 게임콘텐츠스쿨,

\*청강문화산업대학교 게임콘텐츠스쿨

e-mail: {redcom, one, greensu}@ck.ac.kr<sup>o\*</sup>

## An Implementation of Metaverse Virtual Fitting Technology using a Posture extraction based on Deep Learning.

Bum-Ro Lee<sup>o</sup>, Sang-Won Lee\*, Soo-Jin Shin\*

<sup>o</sup>School of Game, ChungKang College of Cultural Industries,

\*School of Game, ChungKang College of Cultural Industries

### ● 요약 ●

본 논문에서는 메타버스 공간에서 패션 아이템 판매에 있어서 필수적이라 할 수 있는 온라인 가상 피팅 기술을 동작 인식 전용 디바이스가 아닌 일반 스마트폰 카메라를 활용하여 구현하는 기술을 제안한다. 가상 피팅 기술을 구현하기 위해서는 딥러닝 기법을 활용하여 입력 영상을 분석하고, 분석 결과를 토대로 인체의 전체 자세를 추정하며, 인체 사이즈의 근사값을 추출하는 과정들이 수행되어야 하는데, 현재의 스마트폰 컴퓨팅 환경은 이를 수행하기에 충분한 연산 성능을 가지지 못한다는 문제점을 가진다. 본 논문에서는 높은 비용이 요구되는 고부하 연산을 클라우드 서버를 통해 수행하는 서버 기반 프레임워크를 도입하여, 낮은 성능의 스마트폰으로도 고성능 연산이 가능한 서비스 구조를 확보하고 이를 통해 휴대성 높은 증강현실 기반의 가상 피팅 기술을 구현한다. 본 논문의 성과를 통해 메타버스 상거래의 활성화와 메타버스 본연의 의미에 충실한 가상 월드 구축에 기여할 것이라 기대한다.

**키워드:** 메타버스(Metaverse), 증강현실(Augmented Reality), 클라우드 컴퓨팅(Cloud Computing)

### I. Introduction

인간의 오프라인 활동을 극단적으로 제한시키는 코로나19의 창궐은 역설적으로 그동안 여러 가지 이유로 미뤄 왔던 온라인 시스템의 강제적인 도입을 촉발시키는 단초를 제공하였고, 팬데믹 상황이 2년 이상 지속되고 있는 현재에는 이러한 움직임을 넘어 메타버스라 명명된 온라인 second life를 구축하고자 하는 방향으로 발전하기에 이르렀다. 현재 가장 주목받고 있는 미래 핵심 키워드인 메타버스는 다양한 해석과 전망에도 불구하고 향후 인간의 생활에 밀접하게 관여하는 차세대 플랫폼으로 자리 잡을 것이라는 의견들이 주를 이루고 있고 해당 분야에 대한 연구와 투자가 지속적으로 증가하고 있다.

메타버스 플랫폼의 가장 중요한 핵심 요소 중에 하나는 온라인에 구현되는 실물 경제 시스템이라 할 수 있다. 이는 온라인의 가상 재화의 거래 뿐 만 아니라 오프라인의 실제 거래를 연동할 수 있는 기능을 포괄하는 개념으로 이를 구현하기 위해서는 다양한 신기술이 요구된다. 본 논문에서는 거대한 오프라인 시장을 형성하고 있는 패션 관련 제품들을 온라인에서 자유롭게 유통하기 위해 필수적인 가상 피팅 기술을 추가적인 고가의 동작 인식 전용 센서 없이 일반적인

스마트폰의 카메라를 활용하여 구현하는 기법을 제시하고 이를 실험을 통해 실증 한다.

### II. Preliminaries

#### 1. Related works

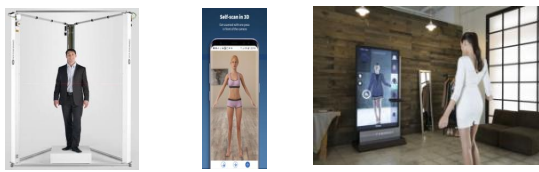
##### 1.1 국내외 기술 동향

다양한 AR기술의 적용 분야 중에서도 가상 피팅 분야는 오프라인 전통 산업과 온라인을 연결하는 매우 중요한 기술로 인식되어 국내외의 연구자들에 의해 다양한 시도들이 진행되어 왔다. AR기술을 활용한 가상 피팅 기술은 크게 인체 영상을 기반으로 자세를 추출하는 단계와 이렇게 추출된 자세 정보를 기반으로 skeleton을 생성하는 단계, 그리고 생성된 skeleton에 피팅 하고자 하는 의상의 3D모델을 추가하여 렌더링 하는 단계로 나누어 볼 수 있는데, 첫 번째 자세 추출 단계의 연산량이 상대적으로 높아 대부분의 기본 서비스에서는 이

단계에서 전용 하드웨어를 활용하는 것이 대부분이다.

Vitronic사는 전신 3D 촬영 기기 제작 업체로서 매장에 설치된 3D 바디 스캐너를 이용한 의류 사이즈 측정 정보 제공한다. 이러한 방식의 서비스는 거대한 실측 기계를 매장 안에 설치해야만 한다는 단점을 가지고 있으며 이로 인해 대중적인 상업화가 어렵다는 한계를 가지고 있다.

Nettelo 3D사의 body scanning and analysis 서비스는 단순한 전방 촬영을 통해 수동적으로 객체 추출하는 방식을 활용한다. 수동 체형 입력 정보를 미리 결정된 Rule-based Size Matrix extraction 기술에 적용하여 신체이치를 측정하는 방식인데 생성 모델 정확도가 낮아서 실용적 수준의 fitting을 구현하지 못하는 문제점을 가진다.



(a) Vitronic (b)Nettelo 3D (c) FxGear

Fig. 1. Existing Services

FXGear의 가상 피팅 시스템은 가장 현대적인 가상 피팅 시스템의 모델을 제시하였으며, 사전 제작된 의상의 모델을 PC에 장착된 Kinect 센서의 입력값을 기반으로 렌더링하고 이를 실사와 합성하여 AR 기반 가상 피팅 서비스를 구현하였다. 그러나 PC 기반의 전용 센서가 필요하다는 측면에서 개인화된 서비스가 어렵고, 모바일 서비스가 불가능하다는 단점을 가지고 있는 서비스라고 할 수 있다.

### 1.2 관련 연구 분석을 기반으로 한 시스템 제안

본 논문에서는 기존 연구 및 상용화 결과를 분석하고 개선점을 도출하여 사용자 별도의 동작 인식 센서나 PC와 같은 고성능 하드웨어의 개입 없이 스마트폰을 활용하여 본인의 신체 사이즈를 측정하고 이를 기반으로 AR 가상 피팅이 가능한 서비스를 설계하고 구현하였다. 이를 위해서 대부분의 고부하 연산을 5G망을 통한 서버 기반 연산으로 처리하고 Deep learning 기반의 자세 및 사이즈 인식 알고리즘을 처리함으로써 스마트폰 단말기의 성능에 관계없이 AR 기반 가상 피팅을 구현할 수 있도록 시스템을 설계 하였다.

## III. The Proposed Scheme

본 논문에서는 신체 실측을 통해 데이터를 추출하며, 신체 데이터를 딥러닝으로 학습 시킨 서버에서 사용자의 체형과 가장 가까운 골격을 추출하고 추출된 골격을 기반으로 제공되는 의상을 피팅할 수 있는 기술을 개발하여 해당 제품 피팅감에 가장 흡사한 Visual Feedback을 느낄 수 있도록 하는데 그 목적이 있다. 또한 필요한 충분한 연산 능력을 가지지 못한 스마트폰 상에서 이러한 기능을 구현하기 위해 소스 영상 획득과 결과 영상 플레이 이외의 모든 과정을 클라우드에 구축된 서버에서 처리하여 기술의 활용성을 높이 방향으로 시스템을 설계 하였다.

## 1. 카메라를 통한 인체 데이터 측정

### 1.1 인체 관절 정보 추출 기술 개발

본 논문에서는 Mask R-CNN 기반의 인체 관절 위치 추출 기술을 활용하여 인체의 자세와 사이즈를 결정하는 기반 데이터를 확보하였다. 스마트폰 카메라를 통해 촬영된 인체를 포함한 영상은 서버로 전송되어 Fig.2와 같은 과정을 거쳐 인체 관절의 위치 정보가 추출되도록 구성하였으며, 획득된 관절의 위치를 연결하여 인체의 전체 자세를 계산하는 방식을 적용하여 가상 피팅을 위한 기본 포즈를 확보하였다.

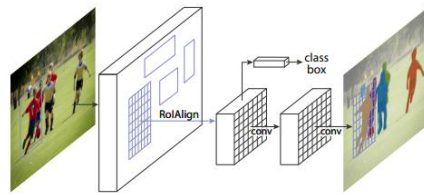


Fig. 2. Mask R-CNN Framework [1]

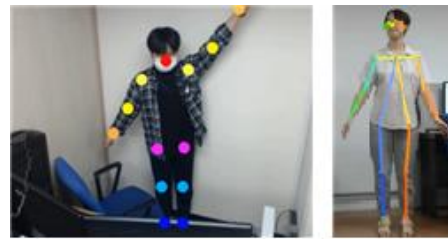


Fig. 3. Extraction of the Human Body Joint Position

### 1.2 카메라 영상 기반 실측 수치 추정 기술 개발

연산의 필요한 깊이 정보는 스마트폰에 탑재된 ToF 센서를 통해 확보하며, 이를 기반으로 카메라 이미지 상의 관절 위치 좌표를 실제 수치로 변환한다. ToF 카메라는 Android SDK의 Camera2 API를 활용하여 접근 가능하며 특정 스마트폰에서는 접근 불가능도록 설정되어 있어 해당 기능을 활용할 수 없는 경우가 발생하기도 한다. 이 경우에는 사용자의 키를 수동으로 입력받아 역산하여 실제 수치를 측정할 수 있도록 하였다.

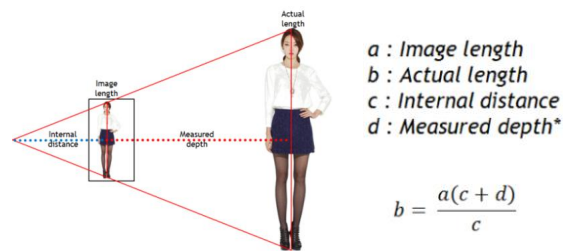


Fig. 4. Calculation for Depth from Plane Image

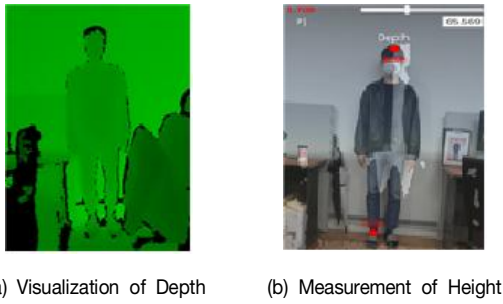


Fig. 5. Calculation for Size by Depth Information

### 1.3 인체 데이터 측정 기술 개발

위 과정을 통해서 추출된 관절의 위치와 깊이 정보만으로 체형에 따른 신체부위의 두께, 둘레를 추정할 수 없기 때문에 세그멘테이션을 이용하여 각 신체부위의 영역으로 측정하여 치수를 측정하는 기술을 구현하였다. 이 과정에서 입력된 신장, 체중 정보와 인공지능을 통해 검출된 각 관절 정보를 활용하였으며 사용자의 실제 신체치수를 추정하는 결과치를 얻을 수 있었다.



Fig. 6. Calculation of Body Size Using Segmentation

### 1.4 AR 의상 합성 기술 개발

본 논문을 통해 구현된 인체 관절정보 및 세그멘테이션을 통한 신체부위 실측 결과를 활용하여 실시간으로 자세와 체형에 맞는 AR 기반 가상 피팅 기술을 개발하였다. 추정된 관절의 위치와 관절간의 거리 값을 활용하여 인체 캐릭터의 Skeleton을 추정하였으며, 이렇게 추정된 인체 Skeleton의 애니메이션 값을 미리 리깁된 의상 모델에 적용하고, 이 결과값을 원 이미지에 합성시켜 AR 가상 피팅 영상을 실시간으로 완성하였다. 구현된 서버 기반 AR 가상 피팅 기술의 시연 이미지를 Fig. 7에 나타내었다.

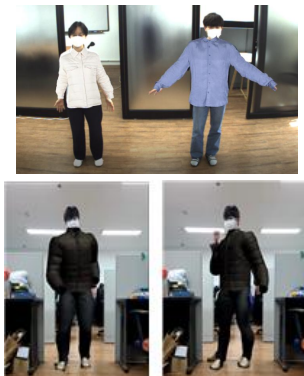


Fig. 7. 3D AR Virtual Fitting Simulation

## 2. 5G 클라우드 서버 구조설계 및 개발진행

본 논문에서는 스마트폰을 통한 고부하 딥러닝 연산의 수행을 위해 데이터 분산처리 구조 및 테스트 베드를 구축하여 이를 기반으로 결과 데이터를 도출하였다. 급격한 계산량 증가로 인한 서버단에서의 scalability에 대한 대응은 복수 서버를 운영을 통한 트래픽 분산 처리로 해결하였으며, 이를 위해 클라우드 서비스에서 제공하는 scale up 기능을 활용하였다. 클라우드 서버의 하드웨어의 특징인 저시양의 다수 CPU 코어 및 GPU 구조를 고려하여 기존 딥러닝 연산 서버의 프로세스 구조를 단일 신경망을 다중 코어로 처리하는 기존 구조에서 신경망을 다중으로 설정하고 각 신경망에 단일 코어가 할당되어 처리하는 구조로 변경하였다. 변경된 구조는 기존 딥러닝 서버 구조와 비교했을 때, 효율적인 분산 처리가 수행되었고, CPU, RAM, GPU 등의 자원이 추가되었을 때 확장 용이성과 하드웨어 자원 활용성이 대폭 증가했으며, 처리 속도도 대폭 증가되는 결과가 도출 되었다.

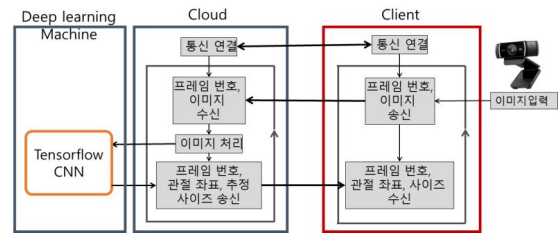


Fig. 8. Server Based Virtual Fitting System

## IV. Conclusions

현재 산업계에서 주목 받고 있는 메타버스 플랫폼은 단순히 온라인에서 사용자들이 모여서 친목 하는 기존의 서비스 범위를 넘어 실제 세계에 대응하는 대부분의 활동이 가능한 확장적 개념을 가지고 있으며, 이러한 측면에서 메타버스의 가장 중요한 요소로 꼽을 수 있는 원활한 경제적 요소를 지원하는 기술에 대한 수요가 폭발적으로 증가할 것으로 예상된다. 본 논문에서 제안되어 실증된 서버 기반 증강현실 가상 피팅 기술은 이러한 요구들에 부응하는 의미 있는 시도 중 하나가 될 수 있다고 판단되며 메타버스 활성화에 반드시 필요한 실제세계와의 인터페이스 기술 분야로서 지속적인 수요가 발생하라는 분야라 할 수 있다.

## REFERENCES

- [1] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick, "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969.
- [2] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu, "Large Scale Datasets and PredictiveMethods for 3D Human Sensing in Natural

- Environments,” IEEE Trans. on Pattern Analysis and Machine Intelligence (2014),36(7):1325-1339.
- [3] Hongsuk Choi, Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee, “Beyond Static Features for Temporally Consistent 3D Human Pose and Shape from a Video,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021).
- [4] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, William T. Freeman, Rahul Sukthankar, and Cristian Sminchisescu, “Generative 3D Human Shape and Articulated Pose Models,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020), pp. 6183-6192.
- [5] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu, “A Hybrid Analytical-Neural Inverse Kinematics Solution for Human Pose and Shape Estimation,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021).
- [6] Kevin Lin, Lijuan Wang, and Zicheng Liu, “End-to-End Human Pose and Mesh Reconstruction with Transformers,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021).