StyleGAN 딥러닝 기술을 활용한 카메라 기반 캐릭터 생성 및 모션 제어 시스템 개발

이정훈[1], 김주형[1], 신동현[1], 양재형[1], 장문수[1*]
한국폴리텍대학 반도체융합캠퍼스 반도체융합SW과
dlwjdgns0308@naver.com, fkjy132@gmail.com, shindonghyun0124@gmail.com, scholar@mine.tel, avecmschang@kopo.ac.kr

Development of Camera-based Character Creation and Motion Control System using StyleGAN Deep Learning Technology

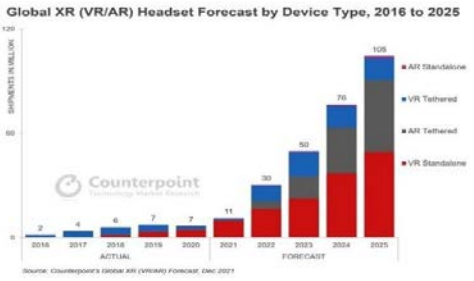
Jeong-Hun Lee[1], Ju-Hyeong Kim[1], Dong-hyeon Shin[1], Jae-hyeong Yang[1], Moon-soo Chang[1*]
[1]Dept. Of Semiconductor Convergence Software, Semiconductor Convergence Campus Korea Polytechnics

요 약

현재 사회적인(COVID-19) 영향으로 메타버스에 대한 수요가 급증하였지만, 메타버스 플랫폼 진입을 지원하는 XR(AR/VR) 장비의 높은 가격대와 전문성 요구로 폭넓은 수요층을 포괄하기 어려운 상황이다. 본 논문에서는 이러한 수요층의 어려움을 개선하고자 웹 캠이나 스마트폰 카메라로 생성된 개인의 사진 이미지를 StyleGAN 딥러닝 기술과 접목시켜 캐릭터를 생성해 Mediapipe를 활용하여 모션 측정 및 제어를 처리하는 서비스를 제안하여 메타버스 시장의 대중화에 기여하고자 한다.

1. 서론

최근 COVID-19의 영향으로 공간적인 제약에 구애받지 않는 메타버스 플랫폼을 통해 사회활동을 하는 인구가 급증하고 있다.



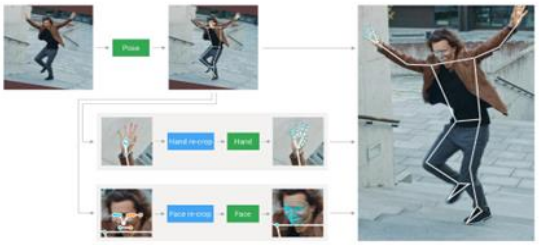
(그림 1) 글로벌 XR 시장 전망

(그림 1)과 같이 2020~2021년부터 비대면 활동의 수요가 급증함에 따라, XR 시장도 빠르게 성장하였고, 애플과 삼성전자와 같은 글로벌 IT 기업들이 시장에 진출할 것으로 예상되면서 시장 규모를 더 키울 것이란 전망이다. **[1]** 하지만 XR 장비의 높은 가격대와 요구되는 전문성이 메타버스 이용에 대한 진입장벽을 높게 형성하여 더 많은 수요층을 포괄하지 못하는 것이 현실이다. 이에 본 논문은 비교적 낮은 가격대로 구입이 가능한 웹 캠과 보유하고 있는 휴대폰 카메라를 활용하여 메타버스 캐릭터의 생성과 모션 인식 및 제어를 처리함으로써 기존 XR 시장대비 낮은 가격층과 간편한 접근성으로 메타버스 사용자층을 증대 시키는 솔루션을 제안한다.

2. 이론적 배경

2.1 Mediapipe

Mediapipe는 구글에서 제공하는 AI 프레임워크로 비디오 형식 데이터를 이용한 다양한 비전 AI 기능을 파이프라인 형태로 손쉽게 사용할 수 있도록 도와준다. 해당 솔루션 중 하나인 Mediapipe Holistic 파이프라인은 각각의 특정 영역에 최적화된 Pose, Face and Hand로 나누어 사람의 모션을 측정할 수 있도록 모델을 통합한 솔루션이다.

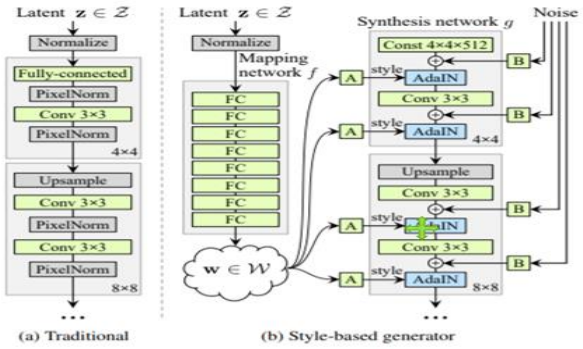


(그림 2) Media Pipe 전체적 파이프라인 개요.

먼저 Blaze Pose 의 포즈 감지기와 후속 랜드마크 모델을 사용하여 (그림 2)의 상단과 같이 인간의 포즈를 추정한다. [2] 그 후, 추론된 포즈 랜드마크를 사용하여 각 손과 얼굴에 대해 세 가지 관심 영역인 ROI(Region of Interest)별로 이미지 포커싱을 유도하고 ROI 를 개선하기 위해 다시 포커싱 모델을 사용하여 정확도를 높인다. 다음 단계로 얼굴 및 손 모델을 적용하여 해당 랜드마크를 추정한다. 마지막으로 모든 랜드마크를 포즈 모델의 랜드마크와 병합하여 전체 540 개 이상의 랜드마크를 생성한다. [2]

2.2 StyleGAN

StyleGAN(A Style-Based Generator Architecture for Generative Adversarial Networks)는 기존 생성적 적대 신경망(GAN, Generative Adversarial Networks)을 고해상도 데이터에 적합한 PGGAN 에서 발전하여 이미지의 특성을 추출하는 데에 좋은 성능을 보이는 NVIDIA 사의 딥러닝 모델이다. (그림 3)과 같이 학습 중간에 기존의 GAN 모델에서 합성곱 계층(Convolution Layer)을 추가해 학습데이터의 선명도를 높이고, 눈, 코, 입, 헤어스타일, 안경과 같은 특정 이미지에 해당하는 특성에 따라 가중치(w: weight)를 각각 추출해 AdaIN(Adaptive Instance Normalization)에 대입시켜 이미지 특성에 대한 학습 성능을 높였다. 이 기술은 각 사람의 얼굴 특성에 기반하여 이미지의 화풍을 변경하는 데에 활용된다. [3] [4]

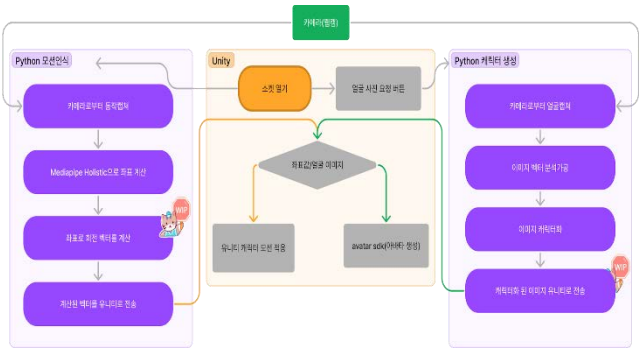


(그림 3) StyleGAN 의 구조

2.3 Avatar SDK

Unity Plugin 인 Avatar SDK 는 딥러닝으로 학습된 모델을 통해 웹 캠으로 입력된 사람의 전신 동영상 데이터를 대상으로 실시간 모션 인식 및 제어를 위한 가상의 캐릭터를 만드는 데에 활용할 Plugin 이다.

3. 구현 절차 및 결과



(그림 4) 서비스 흐름 구성도

본 시스템은 웹캠을 통해 메타버스 캐릭터의 생성 및 제어를 가능하게 한다. 우선, 카메라(웹캠)을 통해 캡쳐된 얼굴 이미지를 소켓통신으로 캐릭터 생성기로 전달한 후, 이미지를 StyleGAN 모델로 가공시킨다. 가공시킨 이미지는 다시 소켓 통신을 통해 Unity 모듈로 전달되어 화풍이 변경된 이미지와 캐릭터를 생성하기 위한 모션 영상을 생성한다.

생성된 모션 영상은 Mediapipe Holistic 을 이용하여 전신 좌표 벡터를 계산한 후 생성된 캐릭터에 적용되어 좌표 값을 실시간으로 적용하여 캐릭터 동작에 따라 신체 각 관절의 좌표 값을 매칭하여 실제 사람의 영상과 캐릭터의 영상이 매칭되어 작동된다.

3.1 캐릭터 생성

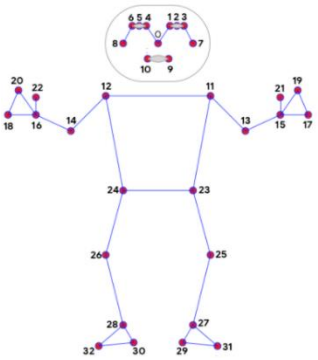
웹캠으로 인식된 사용자의 얼굴 이미지를 분석하여 Latent Vector 값으로 변환 후 모델 데이터 저장을 위해 PT(Pytorch Model)파일로 저장한다. 저장된 파일에서 Style GAN 모델을 통해 특성을 추출하여 인물의 화풍을 변경한다. 위 모델은 특성을 추출하는 데에 특화되어 있기 때문에 화풍 변경을 위해서는 참조할 이미지가 필요하다.

(그림 5) 인물의 화풍 변경

(그림 5)에서는 기존에 특성을 미리 추출하여 저장해 둔 디즈니 화풍 이미지를 참조하여 인물의 화풍을 디즈니 화풍으로 변경하였다.

3.2 Mediapipe Holistic을 이용한 벡터 추출

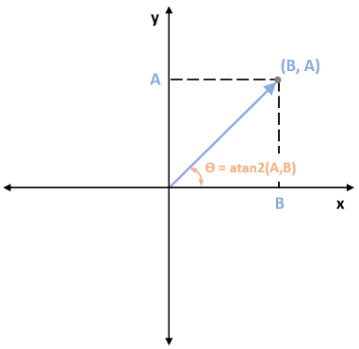
화풍 변경이 이루어진 이미지는 실제 영상의 모션 제어를 위해 (그림 6)과 같은 신체 관절 정보를 참고하여 Mediapipe Holistic을 통해 관절의 좌표를 추출하였다.



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

(그림 6) Mediapipe에서 제공하는 관절 좌표 API

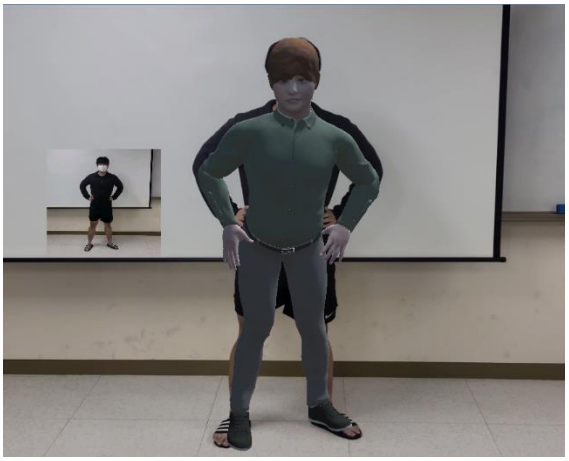
Unity HumanBodyBones API에 맞는 값으로 회전 벡터를 계산하여 동적으로 움직이는 모션에 대한 관절 좌표를 얻기 위해서 회전 벡터를 구하는 공식을 활용하였으며, (그림 7)과 같이 두 점의 벡터를 대입하여 절대 각도를 구하는 atan2 공식을 이용하였다. 예를 들면, 오른쪽 어깨에 대한 회전 벡터를 구하기 위해서 관절 좌표 API 중 12,11,13번 좌표 값을 통해 12->11,11->13 의 두 벡터를 계산한 후 atan 공식을 이용하여 왼쪽 어깨의 절대 각도를 구한다. 이렇게 구한 절대 각도를 StyleGAN에서 생성된 이미지를 적용하고 생성된 Unity 캐릭터를 각 부위에 맞는 HumanBodyBones API로 적용시켜 실제 영상과 모션 좌표가 적용된 영상이 작동되도록 한다. [5] [6]



(그림 7)두 벡터사이의 절대각도를 구하는 atan2 공식

3.3 결과

서버에서 웹 캠을 통해 인물의 얼굴 사진을 전송하여 가공을 거쳐 클라이언트에 전달해주면 클라이언트 측에서 Avatar SDK Plugin을 통해 캐릭터를 생성하고, 생성된 캐릭터를 Media Pipe에서 제공해준 관절 좌표 값을 실시간으로 계산하여 캐릭터에 적용시켜준다.
(그림 8)은 Style GAN과 Avatar SDK를 통해 만든 캐릭터를 Media Pipe Holistic으로 적용시켜 실시간으로 실제 영상과 캐릭터 영상이 생성된 모습이다.



(그림 8) 실제 영상과 생성된 캐릭터 영상 결과

4. 결론 및 향후 연구 방향

본 연구를 통해 메타버스에서 캐릭터를 조작하기 위해서 기존의 높은 가격대와 전문성을 요구하는 XR 장비 대신 상대적으로 낮은 가격대인 웹 카메라를 사용하는 방법에 대한 가능성을 제시해 주었다. 하지만 프로토타입을 제작하여 테스트해본 결과 정적 이미지를 포함한 동영상에서 동작을 캡쳐 했을 때 미세하지만 오차 값이 존재하고, 손으로 얼굴을 가리거나 과격한 동작 등의 행동에서는 캡쳐된 데이터에 오류가 발생하기도 한다. 이런 문제들에 대해서는 향후 연구를 계속 진행하여 해결하는 방안을 마련해야 할 것이다. 이 기술에 대한 연구가 더욱 진행된다면, 메타버스 내에서의 자신에게 맞는 개성 있는 캐릭터로 다양한 사회적 활동을 가능하게 만들 뿐만 아니라, 모션 인식을 통한 정교한 CCTV 분석, 실시간 수화통역 등 여러가지 분야에서 개선할 수 있는 가능성을 제시할 수 있을 것이라 예상된다.

참고문헌

[1]Counterpoint “Counterpoint’s Global XR (VR/AR) Forecast, Dec 2021”
[2]https://google.github.io/mediapipe/solutions/holistic.html
[3]Ian J. Goodfellow, Generative Adversarial Nets(NIPS 2014)
[4]Tero Karras, Samuli Laine, Timo Aila, A Style-Based Generator Architecture for Generative Adversarial Networks (CVPR 2019)
[5]이용준, 김태영 “딥러닝 기반 포즈 인식 및 교정을 통한 효율적인 홈 트레이닝 시스템 개발, 2021”
[6] https://google.github.io/mediapipe/solutions/pose.html

※ 본 프로젝트는 과학기술정보통신부 정보통신창의 인재양성사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.