

포즈 추정을 통한 3D 휴먼 모델의 애니메이션 구현

*장예원 **박병서 **박정탁 **이 솔 ***서영호

광운대학교

*jyw7320@naver.com **bspark@kw.ac.kr **jtpark@kw.ac.kr **solee@kw.ac.kr

***yhseo@kw.ac.kr

Implementation of animation of 3D human model through pose estimation

*Jang, Ye-Won **Park, Byung-Seo **Park, Jung-Tak **Lee, Sol ***Seo, Young-Ho

Kwangwoon University

요약

본 논문에서는 RGB-D 카메라와 Mediapipe 모듈을 이용한 신체 추적 및 리깅 프레임 워크를 제안한다. Openpose 및 Mediapipe를 통해 스켈레톤 정보를 추출할 수 있으며, 이 정보를 그래픽스 엔진의 입력으로 사용하여 휴머노이드 아바타 기능을 통해 각 캐릭터의 아바타가 다르더라도 리깅을 구현할 수 있다. 결과적으로 수작업을 통해 리깅을 구현하는 시간을 단축시킬 수 있다. 두 모듈과 RGB-D 카메라를 통해 획득한 3차원 스켈레톤 정보를 통해 실시간으로 사용자를 추적하고 자동 rigging하는 그래픽스 엔진 프레임 워크를 제안한다.

1. 서론

최근 3차원 그래픽스 기술에 대한 관심이 높아짐에 따라 가상현실(Virtual Reality, VR), 증강현실(Augmented Reality, AR), 혼합현실(Mixed Reality, MR) 기술 또한 발전하는 시대가 열리고 있다.

본 연구는 VR 혹은 AR 콘텐츠 생성을 위해 RGB-D 카메라와 같은 장비를 통해 획득한 3차원 모델을 사용자의 움직임을 통해 조작하는 실시간 자동 rigging 방법을 제안한다. 사용자의 움직임은 Pose estimation을 통해 사람의 신체 관절을 추정하여 이루어진다.

Pose estimation은 사람이나 사물의 위치를 예측 및 추적하는 컴퓨터 비전 기술이다. 사람의 Pose estimation은 얼굴, 신체, 손 등을 대상으로 한다. 얼굴 추정은 얼굴의 구성 요소들을 인식하며 신체 추정과 손 추정은 관절의 위치를 추정한다. 결과 데이터의 종류에 따라 2D pose estimation과 3D pose estimation으로 나눌 수 있으며 본 논문에서는 2D pose estimation을 통해 획득한 결과를 RGB-D 카메라를 통해 획득한 깊이 정보를 통해 3차원 관절 정보로 변환하여 사용한다.

본 논문에서는 Azure Kinect 카메라와 Openpose 모듈을 통해 사용자의 스켈레톤 정보를 획득하고 Mediapipe 모듈을 사용해 사용자의 손, 얼굴에 대한 정보를 획득하고 3차원 정보로 변환하는 방법을 제안한다. 그리고 획득한 사용자의 정보를 통해 3D 모델을 자동 rigging 하여 실시간 애니메이션하는 방법을 제안한다.

2. 스켈레톤 추출

본 논문에서 제안하는 3차원 스켈레톤 추출 방식은 두 단계로 이루어진다. 먼저 RGB-D 카메라를 통해 RGB 영상과 Depth영상을 획득하고 3차원 스켈레톤 정보를 추출한다. 그리고 추출된 RGB 영상에서 Mediapipe를 통해 손과 얼굴에 대한 2차원 스켈레톤을 추출한다. 그리고 추출한 깊이 정보를 통해 2차원 스켈레톤 정보를 3차원 스켈레톤 정보로 카메라 투영 행렬의 역행렬을 사용해 변환한다. 그림 2에서 전체적인 과정을 나타내고 있다.



그림 1. 획득한 Depth 영상

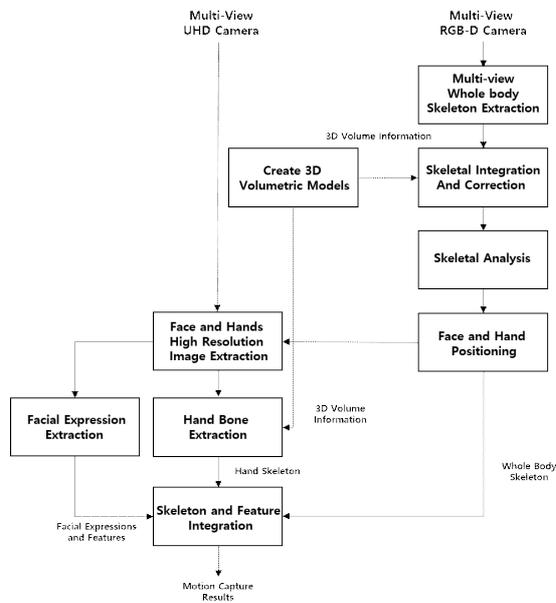


그림 2. 통합적 스켈레톤 추정 과정

먼저, RGB-D 카메라를 통해 촬영된 사용자의 스켈레톤은 Openpose를 통해 추출할 수 있으며 손과 얼굴의 스켈레톤은 Mediapipe를 통해 추출할 수 있다.[1][2] RGB 영상에서 추출된 각 2차원 영상 좌표계 스켈레톤은 투영 행렬의 역행렬을 적용하여 3차원 공간상의 좌표로 변환한다.[3] 이 과정은 식 (1)로 정의된다.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = K^{-1} s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}^{-1} s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (1)$$

일반적으로 2차원 영상 좌표로부터 3차원 좌표를 계산하는 것은 1대의 카메라로 어려운 과정이나, RGB-D 카메라를 통해 얻은 Z 좌표를 통해 유일한 3차원 좌표를 계산할 수 있다.

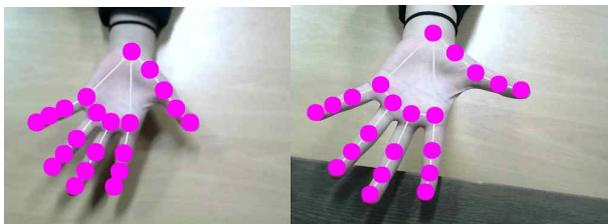


그림3. 손가락 스켈레톤 추출 결과

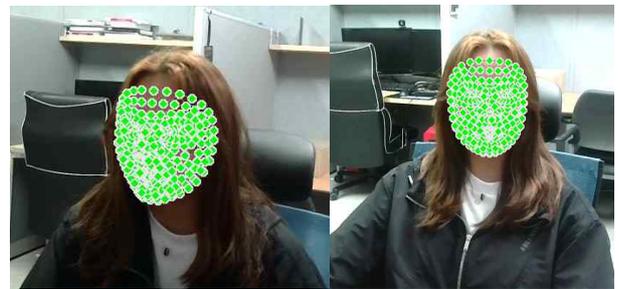


그림4. 얼굴 추출 결과

3. 결론

본 실험을 통해 모션 캡처 센서와 같은 기기를 사용하지 않고 사용자의 3차원 스켈레톤 정보를 추출하여 3차원 모델을 자동 rigging하여 실시간 애니메이션할 수 있는 방법을 제안했다. 결과 영상을 통해 2차원 스켈레톤을 추출할 수 있음을 보였고 식을 통해 이를 3차원으로 투영할 수 있음을 보였다. 추출된 스켈레톤 정보는 3차원 모델 실시간 애니메이션 및 사용자 3차원 관절 정보를 필요로 하는 애플리케이션에 적용될 수 있다.

감사의 글

이 논문은 2021년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 (S2949268).

참고문헌

- [1] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," arXiv, arXiv:1906.08172, Jun. 2019. doi: 10.48550/arXiv.1906.08172.
- [2] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," arXiv, arXiv:1812.08008, May 2019. doi: 10.48550/arXiv.1812.08008.
- [3] 김경진, 박병서, 김동욱, and 서영호. "다시점 카메라 시스템을 이용한 고정밀 스켈레톤 추출 기법." 한국방송미디어공학회 학술대회논문집(2020): 297-299