

## 깊이 영상을 이용한 FBX 형식의 애니메이션 생성 시스템

최진원, \*김상준, \*\*박서연 \*\*곽기석 \*\*박구만  
서울과학기술대학교  
randomkok@seoultech.ac.kr

### FBX Format Animation Generation System using Depth Image

Jin-won Choi \*Sang-joon Kim \*\*Seo-yeon Park \*\*Gi-seok Kwak \*\*Goo-man Park Dept.  
of Mechanical System Design Engineering  
\*Dept. Of Information Technology and Media Engineering, The graduate School of Nano  
IT Design Fusion  
\*\*Dept. of Media IT Engineering  
Seoul National University of Science and Technology

#### 요 약

실시간 동작 애니메이션을 생성하는 기술은 게임, 영화, 의료 등 많은 분야에서 활용할 수 있는 기술이다. 기존에는 마커를 부착하는 모션캡처 방식이 사용되어왔지만 비용, 편의성 등의 문제로 쉽게 사용될 수 없었다. 최근 딥러닝 기술의 발전으로 영상을 통해 사람의 신체 관절 좌표를 추정할 수 있게 되면서 여러가지 방식으로 연구되어 왔다. RGB-D 카메라 등을 통해 얻는 깊이 영상으로부터 3 차원 관절 좌표를 추정하는 방식이 연구되고 있으며, 본 논문에서는 이러한 방식을 통해 실시간 동작 애니메이션을 생성하고 가상공간에 표출하는 시스템을 구현한다.

#### 1. 서론

최근 메타버스와 같은 가상공간의 활용이 커지면서 다양한 콘텐츠를 필요로 하고 있다. 사용자의 동작을 실시간 애니메이션으로 생성하는 기술은 그러한 가상공간 뿐만 아니라 영화, 스포츠, 의료 분야 등에도 제공될 수 있어 기술의 중요도가 높아지고 있다. 사용자의 동작을 애니메이션으로 생성하기 위해서는 각 관절의 3 차원 정보가 필요하다. 3 차원 정보를 얻기 위해서 각 관절에 센서를 부착하는 모션캡처 기술이 사용되어왔다. 하지만 이 방법은 장비의 비용, 장소 문제, 편의성 등의 문제로 쉽게 사용되기 어렵다. 최근 이러한 문제를 해결하기위한 방법으로 딥러닝을 통한 자세 추정 기법이 활발히 연구되고 있다. 이미 2 차원 자세 추정 기법에 대해서는 좋은 성능을 보이고 있다. 하지만 3 차원 자세를 추정하기 위해서는 일반 RGB 카메라로는 어려움이 있어 깊이 정보를 얻을 수 있는 IR 센서, 라이다 센서 등의 사용이 필요하다. 마이크로소프트(Microsoft)에서 개발한 키넥트(Kinect) 센서는 깊이 정보를 얻을 수 있고, 사용자가 쉽게 3 차원 관절 정보를 얻어 게임 및 엔터테인먼트를 즐길 수 있게 되었다. 3 차원 관절 정보를 얻고 애니메이션을 시각화하기위해 전문적인 프로그램을 사용하지 않고도 오토데스크(Autodesk)에서 개발한 FBX 형식을 사용하면 실시간 3D 애니메이션도 구현이 가능해졌다. FBX 형식은 Root

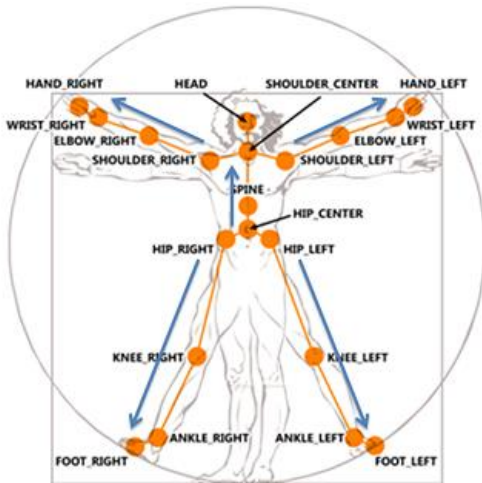
Node 로부터 각 관절을 부모-자식 관계를 정의함으로써 신체 스켈레톤을 구성할 수 있다.

본 논문에서는 키넥트 센서를 이용하여 사용자 신체의 깊이 영상을 얻고, 딥러닝 네트워크를 이용하여 3 차원 관절 좌표를 추정한다. 그리고 FBX 형식의 스켈레톤 애니메이션을 생성하고 가상공간에 표출하는 시스템을 구현한다.

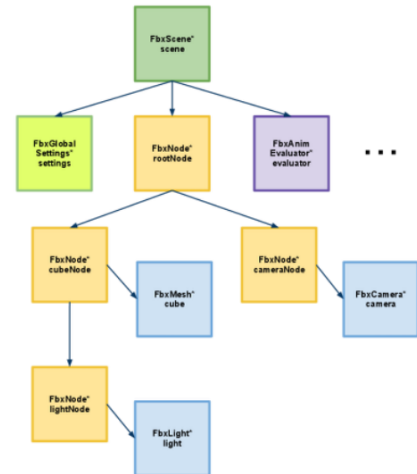
#### 2. 관련연구

##### 2.1 Kinect

마이크로소프트에서 개발한 키넥트는 RGB 카메라와 IR 카메라로 구성되어있다. RGB 카메라는 일반 색상 정보를, IR 카메라는 거리 정보를 얻을 수 있다. 또한 키넥트는 사용자의 25 개의 관절 정보인 Skeleton Joints 정보를 제공해준다. 키넥트가 제공해주는 Joints 정보는 [그림 1]과 같다. 이 정보를 이용해 애니메이션을 생성할 수 있다. 하지만 키넥트가 제공하는 정보의 정확도가 낮아 다른 알고리즘을 사용한다.



[그림 1] 키넥트 스켈레톤 조인트



[그림 2] FBX 구조

### 2.2 3차원 관절 추정

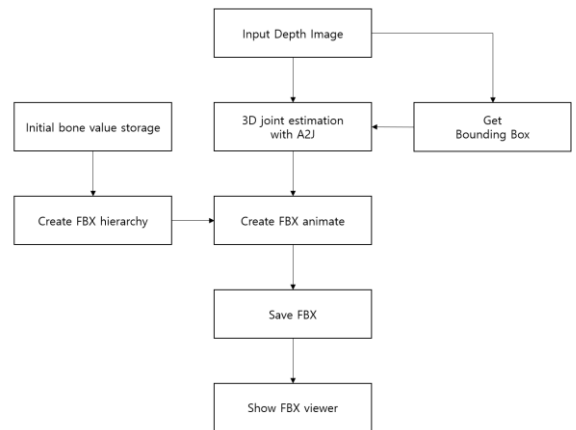
2 차원 관절 추정에서는 DeepPose[1], OpenPose[2]등의 모델이 이미 좋은 성능을 보이고 있다. 3 차원 관절을 추정하기 위해서는 2 차원 관절 추정을 통해 관절 좌표를 얻고, 3 차원으로 변환하는 과정을 거치는 방법과 RGB-D 영상을 통해 3 차원 관절 좌표를 추정하는 방법 등 다양한 방법으로 연구가 진행되고 있다. 이 시스템은 A2J(Anchor-to-joint)[3] 네트워크를 이용한다. A2J 는 깊이 영상으로부터 3 차원 관절 좌표를 추정하는 네트워크이다. Resnet 을 backbone 으로 하여 특징을 추출하고 Anchor branch, Offset branch, Depth Estimation branch 로 구성되어 관절을 추정한다.

### 2.3 FBX 애니메이션 생성

FBX 는 오토데스크가 개발한 파일 포맷이며 오토데스크 Filbox 형식으로 저장되는 3D 모델이다[4]. 다양한 디지털 모델링 및 콘텐츠 생성에 사용할 수 있으며, 게임 개발 및 애니메이션에 자주 사용된다. FBX SDK(Software Development Kit) 를 이용하여 구현 가능하다. FBX 는 [그림 2]와 같이 크게 Manager, Property, Connection, Attribute, Layer 의 개념으로 추상화 된다. 먼저 FbxManager 를 정의하고 FbxScene 을 할당한다. FbxScene 은 노드 계층 구조로 구성된다. Scene 의 요소로 mesh, light, camera 등의 FbxNode 와 FbxNodeAttribute 의 하위클래스와 결합하여 정의된다. 노드는 주로 뼈 내의 뼈 요소의 위치, 회전 및 크기를 지정할 때 사용한다. 실시간 동작 애니메이션 생성을 위해 스켈레톤이 필요하며, 스켈레톤은 Node 를 정의해서 구성한다. 또한 관절들의 부모-자식 관계를 정의하여 골격을 형성한다. 본 논문에서 제안하는 시스템에서는 관절 좌표를 추정한 다음 상대좌표를 계산하여 스켈레톤을 생성하고, 다음 프레임 좌표를 계산, 애니메이션을 생성한다. 그리고 이를 3 차원 가상공간에 표출한다.

### 3. 시스템 개요

본 논문에서는 딥러닝 네트워크를 이용하여 깊이 영상으로부터 3 차원 관절 좌표를 추정하고 애니메이션을 생성, 가상공간에 표출하는 시스템을 제안한다. 시스템 전체 구성도는 [그림 3]과 같다.



[그림 3] 시스템 구성도

사용자의 캐릭터를 생성하기 위해서는 초기 관절 사이의 거리 정보는 고정되어야 한다. 딥러닝 네트워크를 통해 실시간 관절을 추정하여 애니메이션을 생성하면 노이즈로 인해 길이가 달라질 수 있다. 이를 방지하기 위해 이 시스템에서는 Revision Skeleton 정보를 저장한다. Revision 정보는 초기 관절 사이의 거리 정보를 저장하기 위해 수 프레임동안 좌표를 저장하고 평균을 낸 후 길이 정보를 저장한다. 이후 애니메이션을 생성할 때, 각 관절 사이의 방향벡터를 계산한 후 길이를 곱하여 생성한다. 초기 관절 정보를 저장했으면, 실시간 관절 정보를 저장한다. A2J 네트워크를 통해 3 차원 관절 좌표를 얻기 위해서는 사용자의 위치 검출이 먼저 필요하다. 이 시스템은 사용자가 한명으로 가정하여 간단한 알고리즘으로 경계 상자를 생성한다. [그림 4]는 키넥트를 이용해 사용자의 신체 정보를 얻은 영상이다. [그림 5]는 경계 상자를 생성한 영상이다. 사용자가 두 명 이상일때는 객체 인식(Object detection) 알고리즘을 이용하여 각 사용자의 위치를 검출한다. 이후 각 관절의 좌표를 보정하여 저장한다. FBX SDK

스켈레톤을 생성하기위해 저장한 관절 좌표를 불러온다. 주요 관절 15 개의 정보를 저장하였고, 평행 이동(translation)으로만 애니메이션을 구현한다. 애니메이션을 FBX 파일로 저장하고, 가상공간에 표출한다.



[그림 4] 신체 깊이 영상 [그림 5] 경계 박스

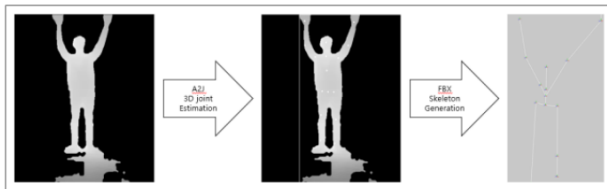
[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7291-7299, 2017

[3] Xiong, Fu, et al. "A2j: Anchor-to-joint regression network for 3d articulated pose estimation from a single depth image." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[4] FBX SDK <https://boycoding.tistory.com/131>

#### 4. 실험

이 시스템은 OpenFrameworks 0.10.0 으로 구현되었으며, Kinect v2 를 이용하여 깊이 영상을 얻었고, Libtorch 라이브러리를 이용하여 딥러닝 모델을 C++에서 이용하였다. 실시간 애니메이션 생성 및 표출은 FBX SDK 를 사용하였다. 실험 결과 깊이 영상으로 관절 좌표를 추정하고 FBX 애니메이션을 생성하였다.



[그림 6] 실험 흐름도

#### 5. 결과

본 논문에서 실시간 애니메이션을 생성하기위해 3 차원관절 좌표를 얻어야 하며, Kinect 에서 얻을 수 있는 Skeleton Joints 로는 정확성이 좋지 않았다. 깊이 영상으로 관절 좌표를 얻을 수 있는 A2J 네트워크를 사용하여 더 정확한 좌표를 얻었으며, 애니메이션 생성을 위해 초기 관절 정보를 이용해 보정하여 더 자연스러운 동작을 구현하였다.

#### 감사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00217, 투명도와 레이어 가변형 실감 사이니지 기술 연구)

#### 6. 참고문헌

[1] Alexander Toshev, Christian Szegedy "DeepPose: Human Pose Estimation via Deep Neural Networks" arXiv preprint arXiv: 1312.4659, 2013