

김세훈 (23학번, 팀장) | 문우혁 (23학번) | 변윤지 (23학번) | Abbasov Nijat (23학번)
이서희 (22학번) | 최재윤 (22학번) | 박철민 (21학번) | 서정근 (21학번)

서론(Introduction)

최근 로봇이 산업계나 사회 전반으로 확장되면서 다양한 일을 수행할 수 있는 로봇들이 주목을 받고 있다. 예를 들어, 물건 집기와 문 열기, 물건 조립 등의 일이 있는데, 이들을 수행하는 로봇의 모션을 구현하기 위해서 강화 학습 알고리즘을 사용하고 있는 경우가 많다. 강화 학습 알고리즘에 경우 로봇 학습에는 좋지만 학습과 최적화에 많은 시간이 소요된다는 문제점을 가지고 있다. 이를 해결하고자 고안한 방식이 실제 사람의 모션을 캡처해 학습에 적용하는 것이다. 모션 캡처 기술을 로봇 학습에 적용한다면 사람과 유사한 움직임이나 현실에 최적화된 움직임을 로봇이 보다 쉽고 빠르게 구현할 수 있을 것으로 기대되어 이 프로젝트를 시작하였다.

이론적 배경(Theoretical Background)

1. Stereo Camera Triangulation

카메라는 CCD등의 센서를 활용해 들어오는 빛을 디지털 데이터로 저장하는 장치이다. 이는 3차원의 정보를 2차원의 평면에 투영하는 기능을 수행한다. 따라서 3차원 정보는 카메라를 통해 저장될 때, 차원 1개의 정보를 잃게 되며 이를 3차원 정보로 복구하기 위해 두개 이상의 카메라를 사용하게 된다. 두개의 카메라를 이용하기 위해서는 카메라 calibration이 선행되어야 한다. 이는 다음과 같다.

① Intrinsic Camera Calibration

카메라는 그 자체의 좌표계를 가진다. 이는 카메라의 픽셀로 표현된다. 카메라 자체의 좌표계를 이미지 좌표계로 변환하고 카메라의 왜곡을 제거하는 것을 intrinsic calibration이라 한다. 즉, Intrinsic calibration은 intrinsic matrix K와 distortion coefficient을 도출하는 것을 목표로 한다. 주로 이는 검정색과 흰색 정사각형이 번갈아 나타나는 checkerboard 이미지를 촬영해 계산한다.

② Extrinsic Camera Calibration

카메라 자체의 좌표계를 이미지 좌표계로 변환한 이후, 두 카메라의 이미지 좌표계 사이의 관계를 도출하는 것이 extrinsic calibration의 역할이다. Extrinsic calibration은 3d rotation matrix R, 3d translation vector C를 구하는 것이다. 이는 두 카메라가 동일한 checkerboard를 촬영하여 구한다.

Calibration이 끝난 두 카메라는 각각 world coordinate로 변환할 수 있으며 이는 식으로 다음과 같이 나타낼 수 있다.

$$x_i \times P_i X = 0 \quad (i = 1 \text{ or } 2)$$

x_i 는 점 카메라에 나타난 좌표, $P_i=KR[i]-C$ 는 camera matrix, X 는 해당 점의 world coordinate이다. 카메라는 이 식을 통해 두개의 연립방정식을 구할 수 있으며, 두개의 카메라를 합쳐 4개의 방정식을 구한다. 4개의 식을 Singular Value Decomposition을 이용해 해를 구해내면 두 이미지로 얻은 좌표를 3차원 좌표로 변환하게 된다.

2. Pose Landmark Detection

사람 이미지에서 관절의 좌표를 구하기 위해 주로 기계학습을 통한 모델이 사용된다. 본 프로젝트에 사용된 MediaPipe의 pose landmark detection 기능 또한 CNN 등의 기계학습 알고리즘을 활용해 사람의 관절 위치를 출력한다.

3. Hardware Composition

Dynamixel AX-12A 모터를 사용하여, 각 팔에 4개의 모터를 장착하여 총 4개의 자유도(DOFs)를 구현했습니다. 전체적으로 두 팔에서 총 8개의 자유도(DOFs)를 구성하였다. U2D2 장치를 사용하여 컴퓨터와의 FTDI통신을 통해 모터를 동작하였으며, 12V 전원 공급을 위해 OpenCM 485 모듈이 장착하였다. 프로그래밍 언어로는 Python이 사용하였다.



그림1-1. Dynamixel AX12-A

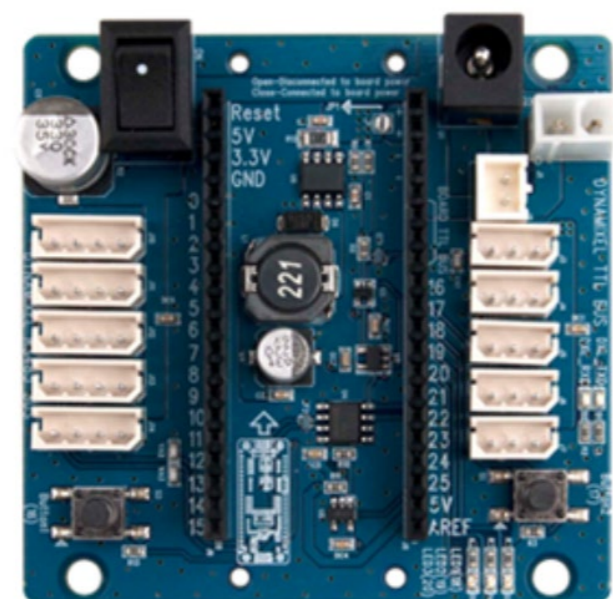


그림1-2. OpenCM 485



그림1-3. Dynamixel AX12-A

제작 과정(Software Methodology)

● 모션 카피 프로그램 (Motion Copying program)

사람의 움직임을 따라하는 로봇을 제작하기 위해선 우선적으로 사람의 모션을 카피할 수 있는 프로그램이 필요하다. 본 프로젝트에선 Google에서 제공하는 Mediapipe 프로그램을 사용했다. 사람의 머리와 사람의 상반신이 카메라상에 잡히게 되면 사람의 주요 관절 부위인 어깨, 팔꿈치, 손목 부분에 점을 생성하게 된다. 결과적으로 카메라 앵글안의 x,y의 좌표를 각각의 점에 대해서 전달해주게 된다.



그림2-1. MediaPipe 로고

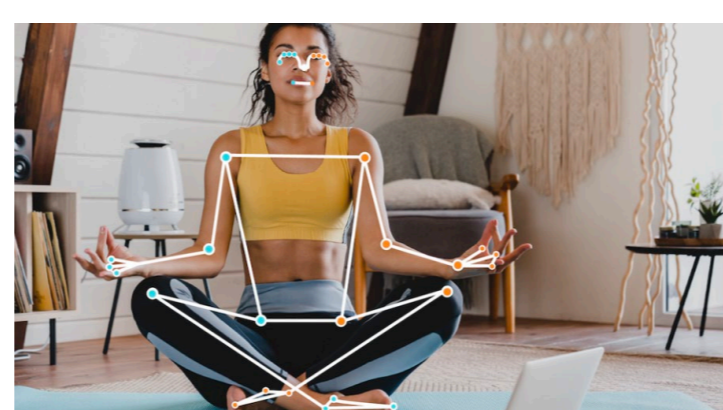


그림2-2. MediaPipe Pose detection

● 캘리브레이션 (Calibration)

본 프로젝트에선 두개의 카메라를 사용하여 3D좌표를 도출하고자 했다. 첫번째로 Checkboard를 사용하여 두 카메라의 캘리브레이션을 진행하고 각 카메라의 matrix를 구하고 두카메라 사이의 관계를 구하기 위해 캘리브레이션을 한번 더 진행했다. 두번째로는 앞서 구한 matrix와 카메라 앵글안의 x, y좌표를 바탕으로 triangulation을 사용해 각 포인트의 3D좌표를 구했다. 다음의 3D좌표는 카메라를 기준으로 하는 좌표계의 좌표를 반환해준다.

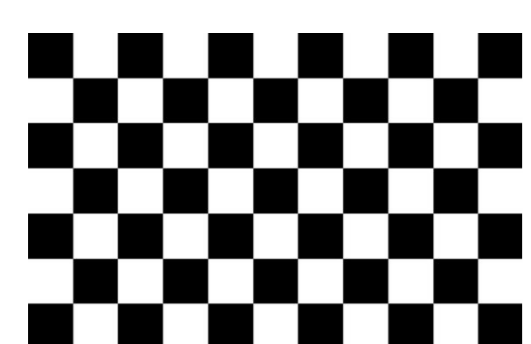


그림2-3. Checkerboard

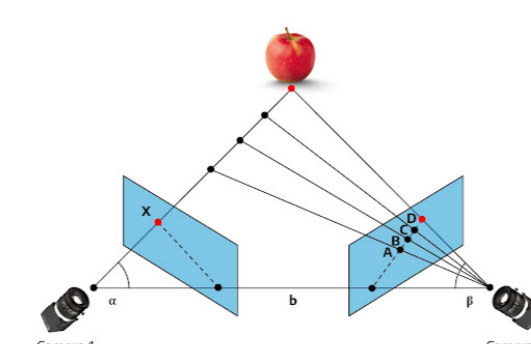


그림2-4. Triangulation

마지막으로 3D 좌표를 바탕으로 로봇에 반환해줄 수 있는 팔의 각도를 구한다. 두가지 부분으로 나눌 수 있는데 첫번째는 어깨부터 팔꿈치까지 윗쪽 팔의 각도이다. 이 부분은 어깨가 고정되어 있기 때문에 위아래 180도 앞뒤 180도로 구성되어 가동 범위를 결정한다. 구하는 과정은 벡터를 사용했다. 두번째는 팔꿈치부터 손목까지 아랫 팔의 각도이다. 이 부분은 윗쪽 팔과 달리 고정되어 있지 않기 때문에 살짝 다른 방법을 사용한다. 우선적으로 윗쪽 팔과 아랫쪽 팔 사이의 각도를 구한다. 다음으로 어깨와 윗쪽 팔 벡터의 외적으로 이루어진 기준 벡터를 바탕으로 아랫쪽 팔의 벡터가 얼마나 회전되었는가를 이야기하는 두번째 각도를 구한다. 결과적으로 윗쪽 팔에선 최대 180도인 각도 2개가 제공되고, 아랫 팔에선 최대 180도인 각도 하나와 최대 360도인 각도 하나가 제공된다. 이것을 모터에 제공하여 사람의 움직임과 동일하게 로봇을 제어한다.



그림 2-5. functions.py

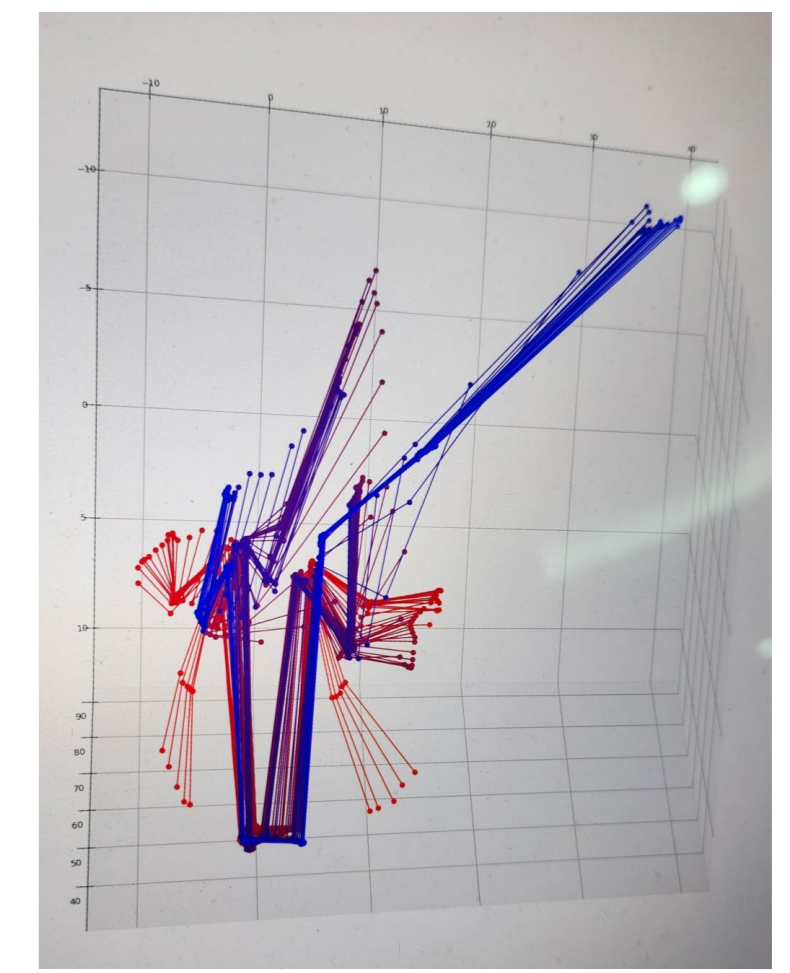


그림 2-6. Pose Detection Result

제작 과정(Hardware Methodology)

하드웨어에서 가장 중요하게 고려한 부분은 로봇의 가동범위가 사람의 관절 가동범위를 모두 포함해야 한다는 것이었다. 사용한 모터 Dynamic Cell의 각도 모드 제어가 가진 한계를 고려하여 다음과 같은 형상의 연결 부품을 설계했으며, 이를 3d 프린터를 통해 제작했다.

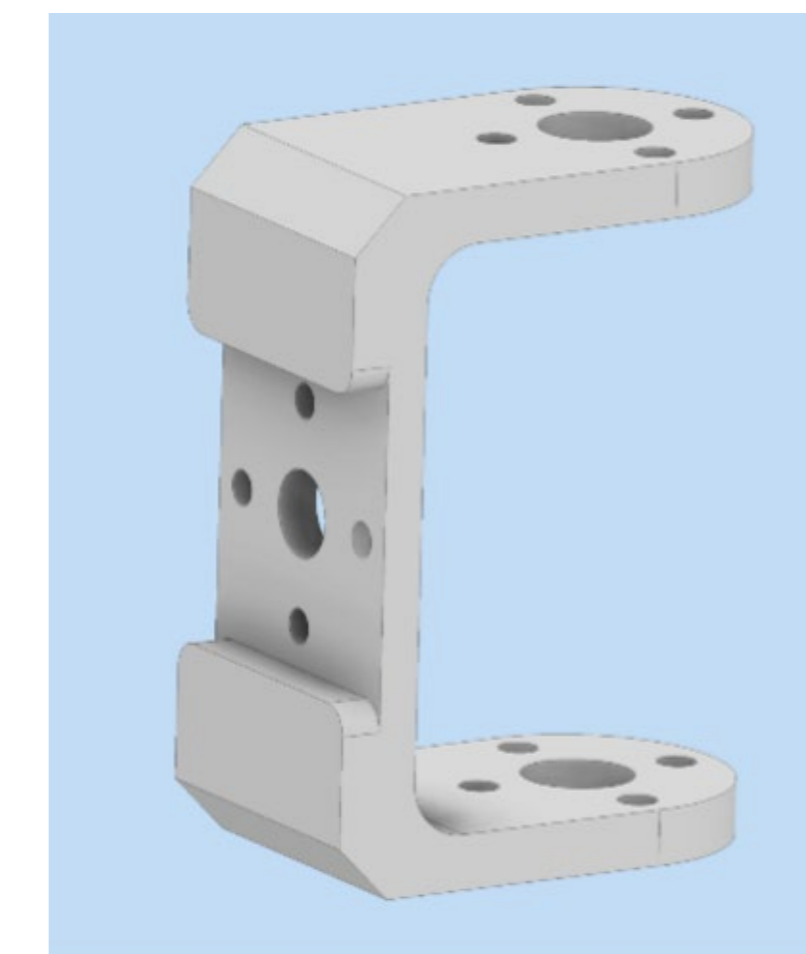


그림 3-1. Joint Connector

제작 결과(Result)

위의 제작 과정을 거쳐서 제작된 Motion Copying Humanoid(MCH)는 캘리브레이션된 2 대의 카메라 앞에 선 사람의 모션을 캡처하고 그 움직임을 거의 유사하게 따로 움직인다. 하드웨어 상의 한계로 사람의 몸과 같은 비율이 아니기에 모션이 사람과 조금은 달라보일 수 있으나, 움직임 자체는 사람의 모션과 정확히 같음을 알 수 있다.



그림 4-1. Motion Copying Humanoid(MCH)

앞으로의 비전 및 활용 가능성(Recommendations)

01

위 기술은 현재 상체만을 구현한 기술로서 추후 연구를 통해서 하체 및 손가락까지 모션 캡처 기술로 구현이 가능할 것으로 기대된다. 또한, 이 기술을 활용해 러닝 프로그램을 제작한다면 머신 러닝을 능가하는 기계 학습이 가능할 것으로 예상된다.

02

이 기술을 실질적으로 응용한다면 사람이 직접 가기에는 위험한 장소에 앞으로 이용될 수 있다. 이에 대한 예시로는 화재 현장에서 사람 구조 및 화재 진압, 붕괴된 건물에서의 탐색, 유독 가스 등으로 접근이 불가능한 장소의 조사 등이 가능할 것이라 예상된다.

03

모션 캡처 휴머노이드는 사람의 행동을 정교하게 따라하여 같은 움직임을 보이는 것이 주된 목표이기에 이에 AI 기술을 적용시키면 전문가의 판단 하에 전문가 이상의 효율을 낼 수 있는 기계의 조합으로 많은 일들을 효율적으로 처리하는데 활용될 수 있을 것이다.