Dry Run (Preliminary) – Virtual Rehab

# PurposE

We aim to establish an efficient and effective workflow for the main data collection for the virtual rehabilitation project through this dry run.

First, we wish to test the chosen *3D camera* and *human/hand pose estimation algorithms* and assess their overall performance throughout the trials:

Some instances for analysis include:

Occlusion, movement artefacts, latency in landmark detection, inaccuracy (visually inspected wrt. RGB-D info)

Second, we wish to receive feedback from the clinicians for experiment design.

i.e. what would make it easier for them to ‘simulate’ patient behavior? do we have representative exercises?

# A picture containing text, floor, person, indoor Description automatically generatedExperiment Set Up

GoPro

GoPro

RealSense D415

## Capture System

1. 1 x Intel RealSense D415 RGB-D Camera [frontal view]

*Resolution*: 1280 x 720 (RGB + D)

*FPS*: 30 (D)

*(output: .bag file)*

1. 2 x GoPro (RGB – wide angle) cameras placed at 45°

## Clinical Exercises

1. Move cup in an arc



1. Move bean bag in an arc



Each clinician will perform the exercises simulating individuals of 4 different categories:

|  |  |  |  |
| --- | --- | --- | --- |
| **1. Mild** | | **2. Moderate** | |
| 1a) High Functioning | 1b) Low Functioning | 2a) High Functioning | 2b) Low Functioning |

\*within each category, we will collect the data from the same clinician three times

\*\*the purpose is to represent variances amongst individuals (thus there will be minor differences between each take

\*\*\* while there were a few other exercises suggested for the dry run, we chose the ones that would be assigned to our entire target population (and only performed two due to time constraints)

# Preliminary Data Collected

A picture containing text, monitor, indoor, electronics

Description automatically generated

**Top**: Circular arc with bean bag exercise with Clinician 1 (left: depth frame sequence, right: RGB video)

**Bottom**: Circular arc with cup exercise with Clinician 2

Output: .bag file

# Further work

* Here we classified patients’ impairment levels in gross categories
  + In order to clarify the categories for clinicians, we must establish an **assessment scale** to use: Fugl-Meyer vs. WMFT-FAS
    - Difficulty here: different publications indicate different ranges to classify as ‘moderate’ vs. ‘mild’
* **Post-process** the output .bag file (MATLAB vs. Python) to get the RGB + depth frames

## Algorithm Selection

As can be seen below, looking at several datasets for human pose as well as hand pose estimation, the top two performing algorithms include:

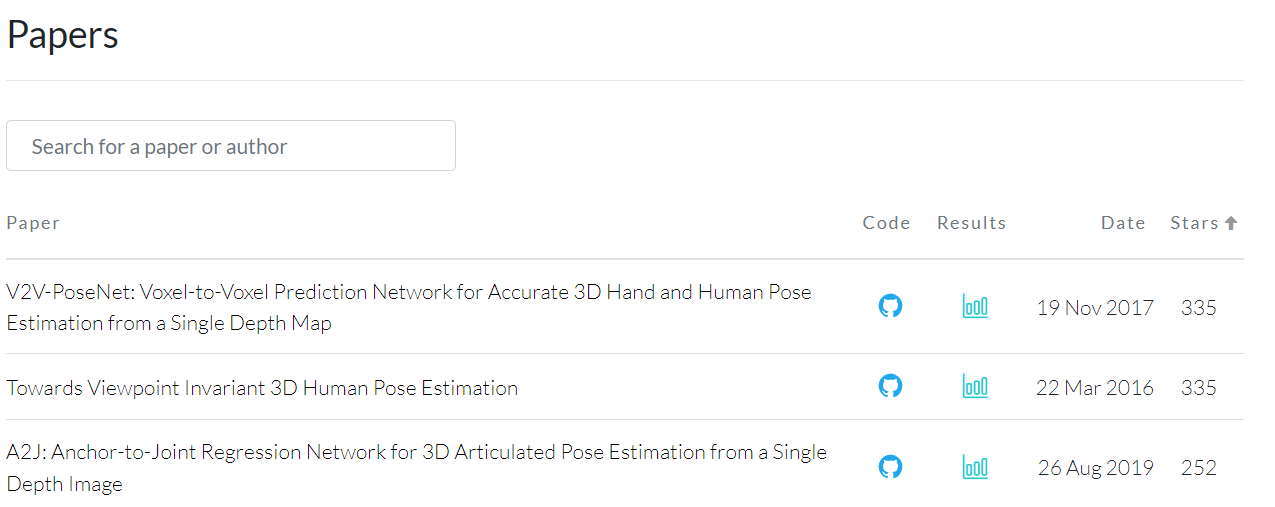
1. V2V-Posenet and 2) A2J: Anchor-to-Joint Regression Network

As such, we will use the above two algorithms for landmark generation.

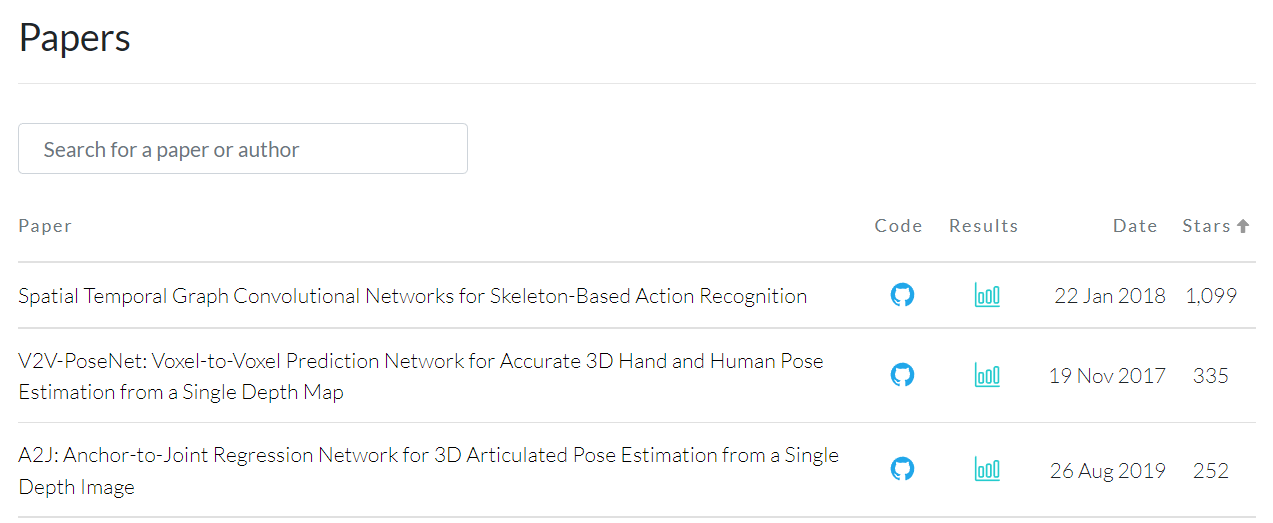
We will compare the performance of the 3D-pose estimate algorithms to the state-of-the-art pose estimation from Mediapipe (from RGB video).

\*we considered comparing it to Openpose (a widely used multi-person pose-estimation algorithm) but according to literature, the performance of Mediapipe is superior

### ITOP (Human pose) Dataset – Papers with Code:



### ICVL (Hand pose) Dataset – Papers with Code:



### NYU HAND (Hand pose) Dataset - Papers with Code:

Graphical user interface, text, application, email

Description automatically generated

## Looking further into the future

We wish to obtain movement characteristics given a video of an individual performing a set of exercises.

Given the landmark positions at different timepoints throughout the video, various motion elements can be obtained.

1. Handcrafted features can be selected and run through traditional machine learning (TML) algorithms – regression, random forest classifier to determine to which category the individual belongs (moderate high, mild low, etc.)
2. Raw data can be input into deep learning algorithms with minimal pre-processing. Consider RNN-LSTM and TCN models that perform well with time-series data

# Appendix

Depth Camera: Intel Realsense D415



|  |  |  |
| --- | --- | --- |
| **Features** | **Use environment:** Indoor/Outdoor  **Image sensor technology:** Rolling Shutter | **Ideal range:** .5 m to 3 m |
| **Depth** | **Depth technology:** Stereoscopic  **Minimum depth distance (Min‑Z) at max resolution:** ~45 cm  **Depth Accuracy:** <2% at 2 m1 | **Depth Field of View (FOV):** 65° × 40°  **Depth output resolution:** Up to 1280 × 720  **Depth frame rate:** Up to 90 fps |
| **RGB** | **RGB frame resolution:** 1920 × 1080  **RGB frame rate:** 30 fps  **RGB sensor technology:** Rolling Shutter | **RGB sensor FOV (H × V):** 69° × 42°  **RGB sensor resolution:** 2 MP |
| **Major Components** | **Camera module:** Intel RealSense Module D415 | **Vision processor board:** Intel RealSense Vision Processor D4 |