Nov 23,2022 Ji Hyun Kim

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:

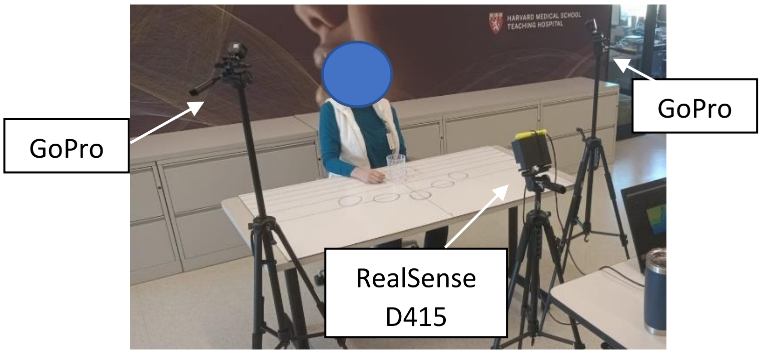
Some instances for analysis include:

> 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?

EXPERIMENT SET UP



CAPTURE SYSTEM

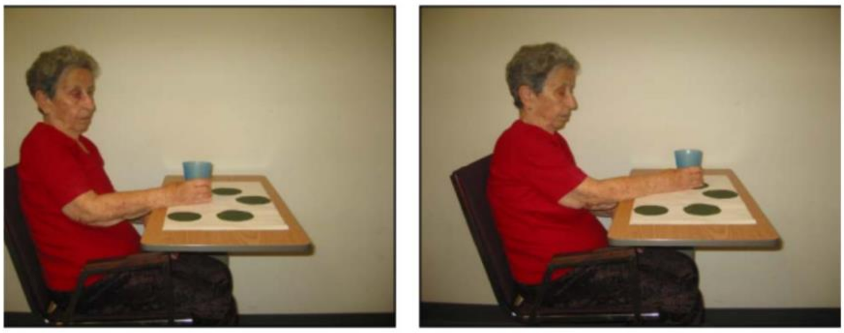
i. 1 x Intel RealSense D415 RGB-D Camera [frontal view]

Resolution: 1280 x 720 (RGB + D)
FPS: 30 (D)
(output: .bag file)

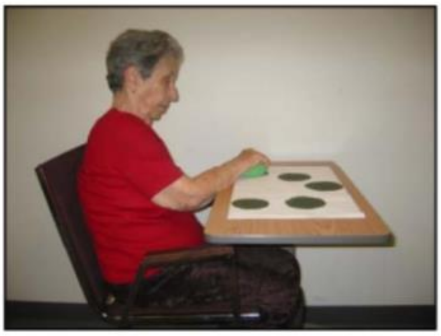
ii. 2 x GoPro (RGB – wide angle) cameras placed at 45°

CLINICAL EXERCISES

i. Move cup in an arc



ii. Move bean bag in an arc



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

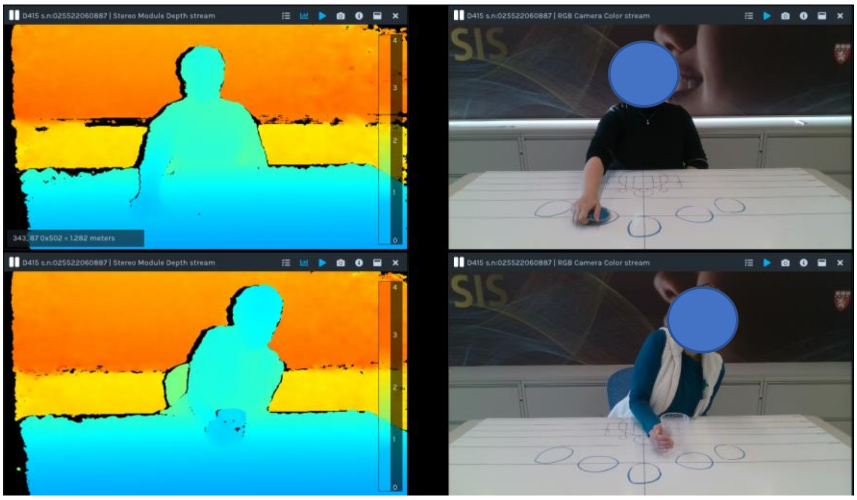
1. Mild		**2. Moderate**	
1a) High Functioning	1b) Low Functioning	2a) High Functioning	2b) Low Functioning

*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


D415 s.n:025522060887 | Stereo Module Depth stream
D415 s.n:025522060887 | RGB Camera Color stream
D415 s.n:025522060887 | Stereo Module Depth stream
D415 s.n:025522060887 | RGB Camera Color stream
SIS
SIS

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:

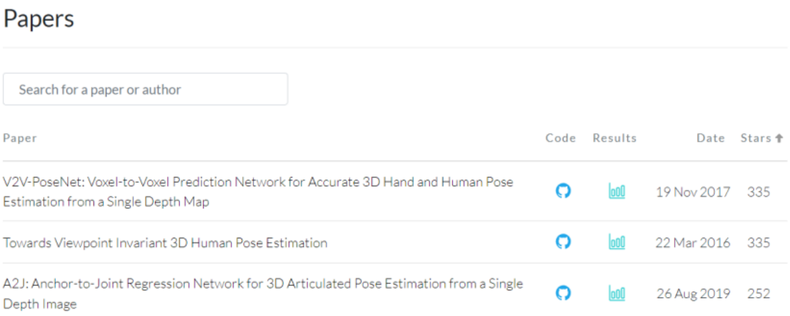
1) V2V-Posenet and 2) A2J: Anchor-to-Joint Regression Network

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:

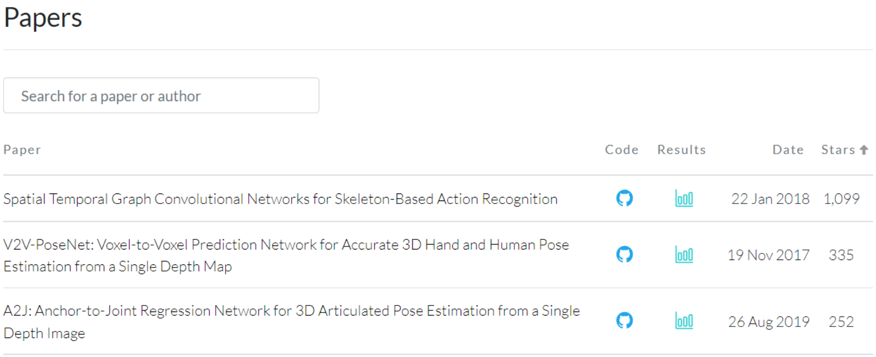


Papers

Search for a paper or author

Paper	Code	Results	Date	Stars ↑
V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map	●	📊	19 Nov 2017	335
Towards Viewpoint Invariant 3D Human Pose Estimation	●	📊	22 Mar 2016	335
A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation from a Single Depth Image	●	📊	26 Aug 2019	252

ICVL (HAND POSE) DATASET – PAPERS WITH CODE:

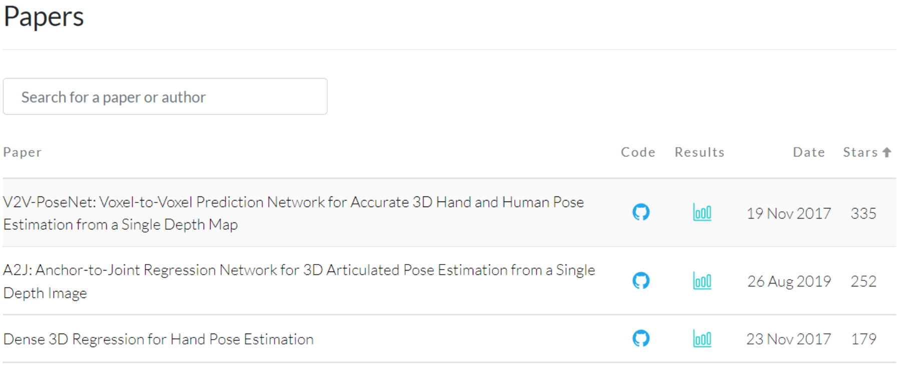


Papers

Search for a paper or author

Paper	Code	Results	Date	Stars ↑
Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition	●	📊	22 Jan 2018	1,099
V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map	●	📊	19 Nov 2017	335
A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation from a Single Depth Image	●	📊	26 Aug 2019	252

NYU HAND (HAND POSE) DATASET - PAPERS WITH CODE:



Papers

Search for a paper or author

Paper	Code	Results	Date	Stars ↑
V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map	●	📊	19 Nov 2017	335
A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation from a Single Depth Image	●	📊	26 Aug 2019	252
Dense 3D Regression for Hand Pose Estimation	●	📊	23 Nov 2017	179

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 m[1]	**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