A comparison of joint degrees between the VICON motion capture system and the MediaPipe Model implemented in Python

Guillem Cornella i Barba & Luis Garcia Fernandez

Abstract—This project aims to compare the performance of two different camera-based approaches to analyze lower-body joint degrees. The comparison is between the VICON motion-capture system and the computer vision-based MediaPipe framework implemented in Python. A GUI has been designed to provide physically impaired individuals with rehabilitation exercises that can be performed from home and exclusively with a smartphone.

Index Terms—Computer Vision, marker-less, telerehabilitation, stroke, VICON, Python, OpenCV, MediaPipe

1 Introduction

Several factors like the aging population, the recent corona pandemic and the need for equally accessible rehabilitation services have increased the demand for telerehabilitation solutions that are cost effective and easy-to-use. [2]

Patients recovering from impairments such as stroke, Spinal Cord Injury, Parkinson's disease or other physical impairments undergo extensive physical rehabilitation and are assessed by their clinicians. This process is usually expensive, time consuming and subject to human error. Furthermore, these kind of assessments may be inaccurate, since they are usually evaluated on visual progress. Integrating assessments that are based on kinematic parameters instead of just using visual evaluations from the clinicians can produce more robust and accurate results. [1]

Computer Vision (CV) modelling and analysis of human motion has been largely studied lately for pose estimation and activity/gesture recognition. These techniques have a huge potential for telerehabilitation services, but the most precise and cutting edge motion capture systems are very expensive and use body worn sensors or are marker-based, meaning that they require the placement of many markers on the body of the subject to be able to track the positions of the joints. This is a very time consuming process and can be intrusive to a patient's day to day activities.

Marker-less CV-based human motion modeling systems using a single camera have the potential to provide home-based, inexpensive and unobtrusive monitoring that can be very useful for telerehabilitation purposes. In [3] a CV-based application employing a webcam is used to measure knee angles, and the results obtained are compared to the measures taken with a universal goniometer.

In this paper, we decided to answer the following question: *Can existing CV-based marker-less human pose estimation techniques, based on a single camera, provide adequate joint localization accuracy for rehabilitation purposes, compared to a marker-based system like VICON?*

A comparison between the measurements taken using the MediaPipe (MP) framework from Google taking images from a single phone camera, and the angles obtained from the VICON motion capture system for the knee and ankle angles is presented. Furthermore, a novel telerehabilitation application is introduced for training and assessing knee flexion and extension.

2 METHODS

To answer our main research question presented in Section 1, we first had to gather data using the two mentioned methods, which are going to be further explained in the following sections.

2.1 MediaPipe Pose Algorithm

Pose estimation is a popular task in Computer Vision, as it enables machines to perform image processing tasks that imitate human vision. By performing pose detection and pose tracking, computers can understand the human body language. Nevertheless, these models must be fast enough and robust enough to occlusions to be viable. High-performing real-time pose detection and tracking will drive some of the biggest trends in computer vision in the following years. Many frameworks have been studied for pose tracking, such as the openPose project from Carnegie Mellon University, HRNet, DeepCut, AlphaPose, DeepPose, PoseNet, DensePose, etc. Nevertheless, MediaPipe was selected as the main framework for this project due to the following factors:

- Better performance indices. The accuracy metrics such as the pck@0.2 are better compared to other frameworks.
- Accurate localization of 33 key points. More than other models.
- Real-time performance on mobile phones with CPU inference. Useful for telerehabilitation.
- Its architecture is available for the implementation of Machine Learning models.

The human pose-estimation pipeline consists of a tracker and a detector, inspired by the *BlazeFace* model. This model has been used inside our Python solution to extract the angles of different joints. The main pipeline can be simplified in Figure 1.

As opposed to the VICON system, which requires 16 markers for a full lower-body human model, the MediaPipe framework uses just 10 lower-body joint markers, hence the resolution can decrease for certain measures.

2.2 Walking and Sitting Tasks

One participant carried out four different tasks to compare the performance of the MP algorithm against the VICON motion capture system. The participant was recorded at once by the VICON cameras and a single phone camera while performing the tasks. The participant was wearing 16 markers on his lower body so that the VICON system could record the joint positions and angles.

The tasks can be divided into two walking tasks and two sitting tasks. The first walking task was normal walking, while the second one was walking lunges (see Figure 2). The first sitting task consisted of flexing and extending the knee while sitting on a stool, and the second one of moving the tip of the foot up and down while keeping the knee straight to produce ankle plantar flexion and dorsiflexion (see Figure 3).

The data obtained from VICON were the coordinates of all the markers and also the joint angles during the whole task. The ground reaction forces measured by the force plates in each belt of the treadmill were also acquired.

For the walking tasks, the vertical ground reaction forces were used to discriminate each gait cycle. Then, the angles for both the knee and ankle were normalized. There was no need to filter the force and marker data since the VICON system had already applied a fourth-order Butterworth low-pass filter with a cutoff frequency of 25Hz.

Regarding the phone-based solution, the joint coordinates were captured with the BlazePose model, and the angles of the joints were computed using vectors with Python functions. A .mat file was generated with the values of the angles obtained from MP and imported into Matlab. That data was filtered using a fourth order Butterworth low-pass filter with a cutoff frequency of 14Hz (the sampling frequency for the MP algorithm was 29Hz).

For the walking tasks, since there was no ground reaction force to discriminate each gait cycle, it was assessed visually by selecting the indices that made the gait curves look more similar to the ones obtained from VICON.

2.3 Lower Limb Rehabilitation Assessment GUI

Once the two data acquisition approaches were well-defined, we designed a GUI to assess lower-limb rehabilitation from home. When designing this experiment, we took into account the needs of physically impaired individuals who suffered from a stroke or spinal cord injuries, since this is the main line of research of the BioRobotics laboratory at UCI, where we are currently completing our Ph.D. studies.

We are proposing a simple platform to offer home-based, inexpensive, and sufficiently accurate exercises, mainly focused on knee and ankle rehabilitation. Such exercises are based on repetitive tasks and instant point rewards, which have been proven to be beneficial (see Figure 4).

3 Results

3.1 Comparison between VICON and MP

3.1.1 Walking Tasks

After analyzing, filtering and normalizing the data from both VICON and MP to a single gait cycle (from heel strike to heel strike), the following plots were obtained.

The plots obtained for both the normal walking and walking lunges tasks for the knee and ankle angles obtained from VICON can be found in Figure 5.

For the MP algorithm, they can be found in Figure 6 for the 3D space and in Figure 7 for the Sagittal plane.

Comparing these three figures to each other is a difficult task. For that reason, they have been grouped into a single figure. In Figure 8, one can find a summary of the three previous figures, were only the averages of each task and joint are plotted.

Looking at Figure 8, it can be observed that the knee angles obtained using MP and a single camera resemble the ones obtained using the VICON motion capture system. Nevertheless, the angles obtained using MP for the ankle do not follow the same pattern as the ones measured with VICON.

The error between the VICON and MP systems has been computed and plotted in Figure 9 as the absolute value of the difference of both MP curves with the VICON curve.

It can be seen that the errors for the knee angles are lower for both walking tasks, rarely being higher than 20º, while the error for the ankle angle can get up to 60º at some points of the gait cycle.

3.1.2 Sitting Tasks

For the sitting tasks, since there was no constant pattern (like the gait cycle for the walking tasks) the whole task was plotted. In Figure 10 one can find the plot for the knee angles in the top and the plot for the ankle angles in the bottom.

From looking at Figure 10, it can be seen, as before, that the values computed by MP for the knee angle are more accurate than the ones for the ankle. MP tracks the movement nicely, even though it does not seem to reach the same values in the peaks.

The error between the VICON and MP systems for the sitting tasks can be found in Figure 11. Once again, the errors for the knee angles are smaller than the ones for the ankle angles.

3.2 Real-time tracking

The GUI presented in section 2.3 allows real-time tracking of several joint angles and is able to return the error between the proposed exercise line (ground truth) and the user's movements detected both by the VICON cameras and the Computer-Vision software. (See Figure 12)

The data captured from the phone's camera is transmitted via UDP-IP protocol to an external plotter with almost real-time accuracy, and the errors are quickly assessed.

The data captured from the VICON cameras is also transmitted via IP protocol. However, in this case the transmission is TCP/IPv4 instead of UDP. The VICON system has an open DataStream SDK to stream in real-time to 3rd party programs, such as our plotter. Their SDK has been modified to send just the essential data, i.e. knee and ankle degrees.

4 Conclusions

After all the tests, we can say that we accomplished some of our proposed objectives:

- The use of a phone camera is cost-effective and there is no need to buy extra equipment.
- Rehabilitation patients do not have to travel to visit physical therapists, avoiding long trips.
- The exercises are easier to integrate into daily activities since they perform the tasks from home.
- There is a real-time evaluation of the performance, with a negligible time delay. Users get instant feedback on their actions.

Nevertheless, as mentioned in [2], the maximum tolerable error margin in joint angle measurements for physiotherapy purposes is ±5°. Therefore, our proposed solution might not be sufficiently precise in some instants to detect the joint degrees for rehabilitation purposes. There are moments when our errors are below 5º, however, when there are occlusions, the error increases as some key points are hidden and the model does not discern the joint coordinates properly. Also, it is possible that the model has not been trained enough with specific side views that we are using for the sitting tasks.

Furthermore, a single camera might not be able to capture the depth information adequately, as we experience larger errors for 3D angle calculations, i.e. using 3D vectors. We need better single-camera 3D pose estimation algorithms to enhance landmark detection and tracking.

We believe that if the environment is better prepared, then the results would be better, i.e. minimizing occlusions, reducing the background obstacles, and placing the camera in an optimal position. The results might be sufficient for simple tasks, yet for more precise exercises, we would need better accuracy and fewer errors.

5 Future research

A proposed solution to enhance depth perception could be to use a depth camera such as the *RealSense* or to access the new LIDAR sensor that is included in the new iPhones and Ipads. Using depth as an additional channel besides RGB frames highly improves classification in these kinds of deep-learning models.

No filtering was performed in this project, so a future solution would include a data filtering step to eliminate disturbances and high errors.

Furthermore, the MediaPipe framework was supposed to be one of the best to fulfill our needs, however, we would like to test other frameworks for the specific rehabilitation tasks that we are performing. Perhaps other models have been trained differently and have different performances for the sitting tasks being evaluated.

References

[1] Bappaditya Debnath et al. *A review of computer vision-based approaches for physical rehabilitation and assessment*. Multimedia Systems, pp.1-31, June 19, 2021.

[2] Thomas Hellsten et al. *The Potential of Computer Vision-Based Marker-Less Human Motion Analysis for Rehabilitation*. Rehabilitation Process and Outcome, 10, p.11795727211022330, 2021.

[3] Thomas Hellsten et al. *Towards Accurate Computer Vision-Based Marker Less Human Joint Localization for Rehabilitation Purposes*. 2022.

6 FIGURES

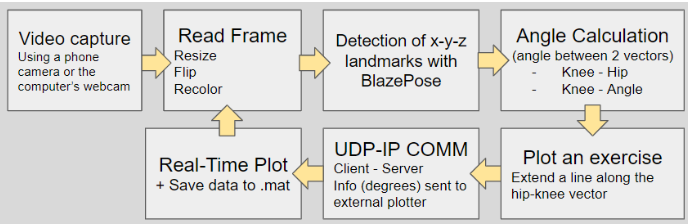

Fig. 1: Code architecture of the Python implementation of the MediaPipe framework for the localization of landmarks/joint coordinates.

Fig. 2: Walking tasks: Normal walking (left) and walking lunges (right)

Fig. 3: Sitting tasks: Knee flexion/extension (left) and ankle plantar flexion/dorsiflexion (right)

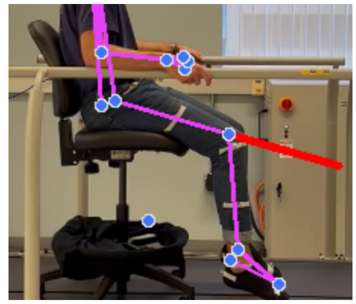

Fig. 4: Proposed exercise for knee-rehabilitation. The user has to move and match the leg with the red line, which keeps changing its orientation after each successful match

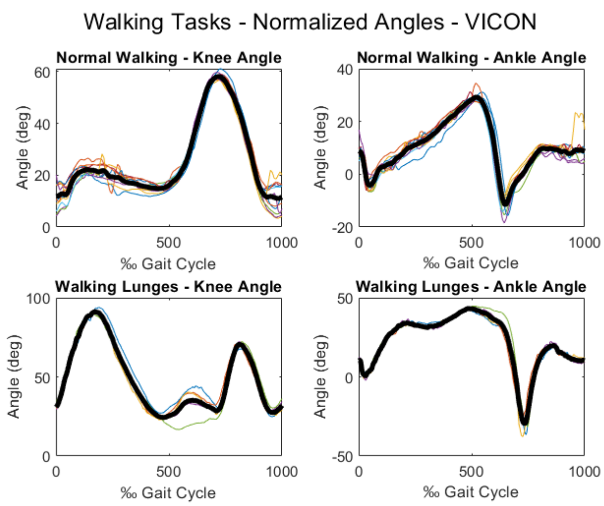

Fig. 5: Walking tasks: Normalized knee and ankle angles for VICON

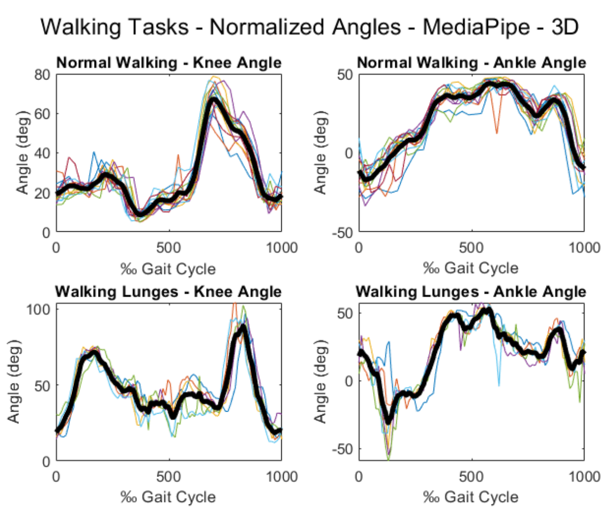

Fig. 6: Walking tasks: Normalized knee and ankle angles for MP in 3D.

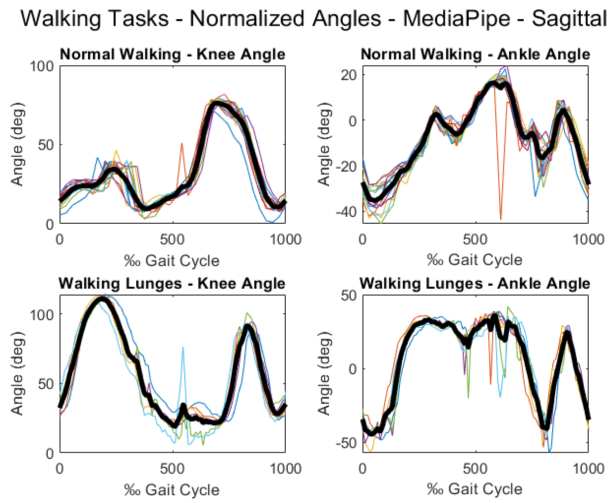

Fig. 7: Walking tasks: Normalized knee and ankle angles for MP in the Sagittal plane

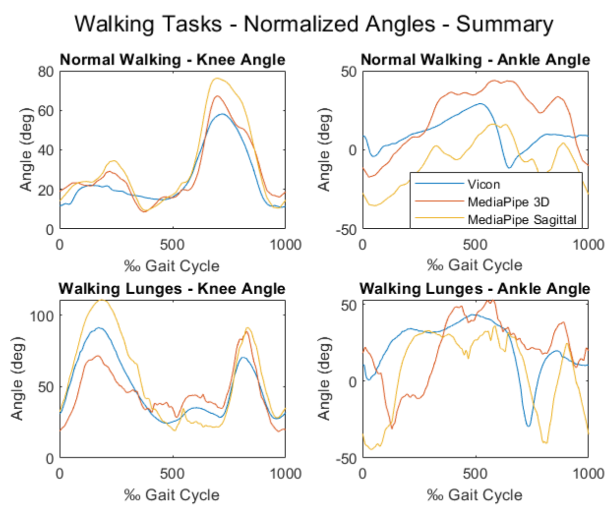

Fig. 8: Walking tasks: Summary

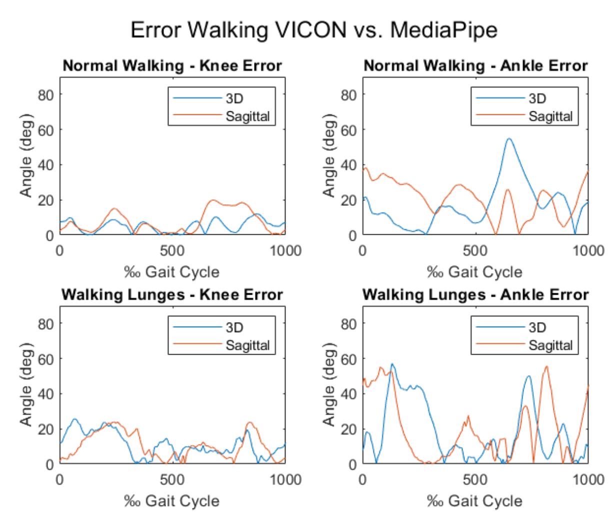

Fig. 9: Walking tasks: Error

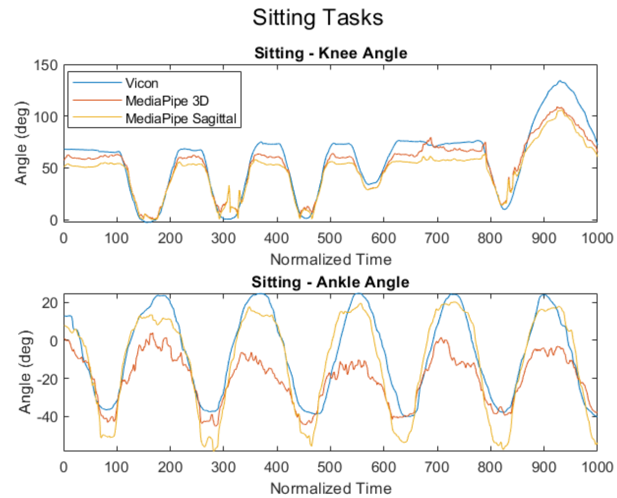

Fig. 10: Sitting tasks: Knee and ankle angles for VICON and MP in 3D and the Sagittal plane.

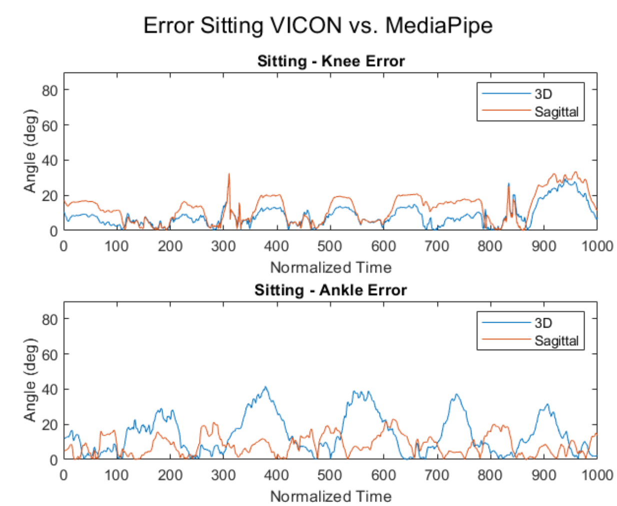

Fig. 11: Sitting tasks: Error

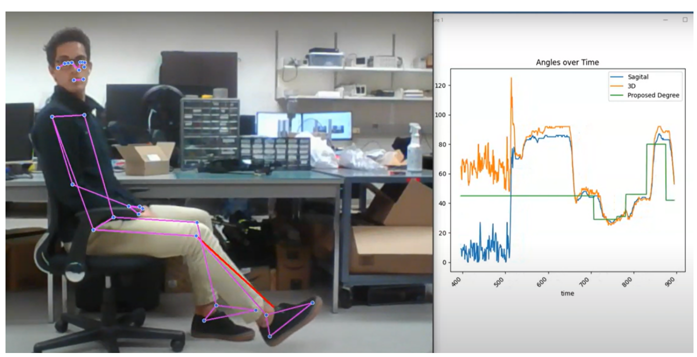

Fig. 12: Real-time tracking experiment with the external plotter.