Accuracy Evaluation of 3D Pose Estimation with MediaPipe Pose for Physical Exercises

S. Dill, A. Rösch, M. Rohr, G. Güney, L. De Witte, E. Schwartz, C. Hoog Antink

BMT 2023

TECHNISCHE UNIVERSITÄT DARMSTADT

Abstract: *With the recent increase in interest in machine learning and computer vision, camera-based pose estimation has emerged as a promising new technology. One of the most popular libraries for camera-based pose estimation is MediaPipe Pose due to its computational efficiency, ease of use, and the fact that it is open-source. However, little work has been performed to establish how accurate the library is and whether it is suitable for usage in, for example, physical therapy. This paper aims to provide an initial assessment of this. We find that the pose estimation is highly dependent on the camera's viewing angle as well as the performed exercise. While high accuracy can be achieved under optimal conditions, the accuracy quickly decreases when the conditions are less favorable.*

Keywords: *Computer Vision, Pose Estimation, Motion Analysis, MediaPipe, Accuracy Evaluation*

Introduction

Libraries like MediaPipe Pose (MPP) [1] (output displayed in Fig. 1) have been applied in diverse fields, including complex medical applications such as joint load prediction [4] and movement limitation analysis [6], indicating potential use in at-home physical therapy supervision.

Problem: This application requires high responsibility and trustworthiness due to its impact on user health. Incorrect executions during at-home exercises could lead to inefficiency or injuries [3].

Goal: This work aims to do a first step towards a quantitative evaluation of MPP, to asses its accuracy when calculating metrics relevant to physical therapy and identify its strengths and weaknesses.

Experiment:

- Pose data of 4 subjects performing physical exercises was gathered using two cameras (C1 and C2) and a motion capture (MoCap) system for ground truth (GT) comparison.
- Five stationary exercises (two push-up variants, squats, kick-backs on all fours, ground swimming) were recorded from different angles.
- Overall, 240 exercises instances were recorded.

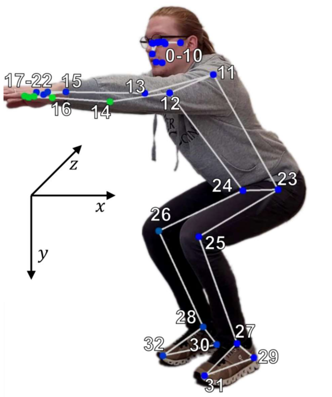

Fig. 1: Visualization of the MPP output of a person performing a squat. The output consists of x-y-z-world coordinates of 33 different landmarks, with the origin lying between the hip joints (23 and 24).

Methods

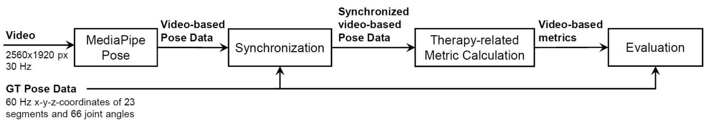

Fig. 2: The data processing pipeline. The GT data was provided by the MTw Awinda MoCap system. For synchronization, a movement metric was calculated along the vertical axis. To focus on raw pose estimation quality, the calculated metrics were consistency of shoulder width (SW) and hip width (HW) as well as flexion angles of knees and elbows.

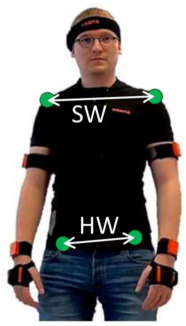

Fig. 3: Visualization of the evaluation metrics shoulder width (SW) and hip width (HW). They were calculated directly form the MP output.

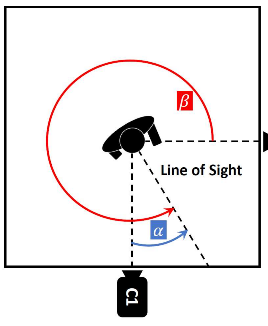

Fig. 4: Top view of the experimental setup. The cameras were placed such that the image plane is vertical to the ground. The camera viewing angles are denoted by α for C1 and β for C2.

Funding

This research is funded by the Hessian Ministry for Digital Strategy and Development [Hessisches Ministerium für Digitale Strategie und Entwicklung, Distr@l-Förderlinie 2, "SG4smartmedication", 21_0038_2A].

HESSEN
Hessische Staatskanzlei
Hessische Ministerin für Digitale Strategie und Entwicklung

digitales.hessen
DISTR@L

Results

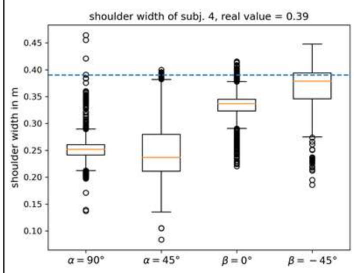

Fig. 5: Boxplot showing the shoulder width of subject 4, estimated for each set (S1/S2) and from both cameras (C1/C2).

Subject (Angles)	SW (C1)	SW (C2)	HW (C1)	HW (C2)
1 ($\alpha = 90°, \beta = 0°$)	15.39%	14.31%	6.24%	16.63%
1 ($\alpha = 180°, \beta = 90°$)	15.15%	11.18%	17.43%	9.78%
2 ($\alpha = 90°, \beta = 0°$)	22.55%	14.30%	4.99%	19.36%
2 ($\alpha = 30°, \beta = -60°$)	20.49%	17.83%	17.30%	30.65%
3 ($\alpha = 90°, \beta = 0°$)	26.41%	10.70%	5.40%	17.89%
3 ($\alpha = -90°, \beta = 180°$)	8.64%	11.27%	34.37%	19.91%
4 ($\alpha = 90°, \beta = 0°$)	29.05%	5.71%	11.90%	24.23%
4 ($\alpha = 45°, \beta = -45°$)	29.78%	11.22%	12.06%	33.81%

Tab. 1: RMSE of the relative SW and HW, as estimated from C1 and C2, separated by set. All values are relative to the normalized measured widths for each subject.

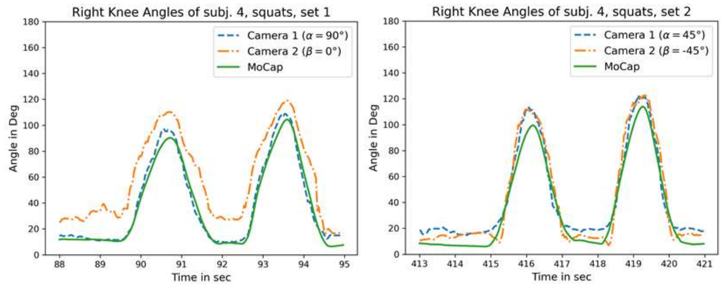

Fig. 6: Plot showing the right knee angles of subject 4 while performing two repetitions of the squat exercise. The first set is shown on the left and the second set is shown on the right.

Exc./Angle	(S1,C1)	(S1,C2)	(S2,C1)	(S2,C2)
Push-Up/RE	15.32	34.96	26.58	23.30
Push-Up v.2/RE	22.40	33.71	29.34	32.71
Squat/RK	9.14	14.19	12.45	14.48
Kick-Back/RK	24.27	18.76	24.90	25.56
Swimming/RE	23.48	50.76	42.22	42.61

Tab. 2: RMSE over all subjects of one important angle for each exercise, for both sets and cameras. RE: right elbow angle, RK: right knee angle. All values in degree (∘).

Discussion

Body Dimension Comparisons:

- Width estimations are biased and not constant due to a lack of subject size information.
- A strong dependence on the camera viewing angle is obvious.
- Lower interquartile range/median ratio for angles close to 0° or 90°, higher for 45°.
- To combat bias, shoulder and hip widths were normalized by shoulder heights and root-mean-square-error (RMSE) calculated for all frames, but the angle dependency is still pronounced.

Angle Estimation Results:

- Viewing angle dependency is noticeable, but less pronounced than for body dimensions.
- Larger deviations from true value seen as angle approaches 0°.
- Angle estimation are most accurate for the squats, where the subjects are standing.

References

[1] Bazarevsky V, Grishchenko I, Raveendran K, et al. BlazePose: On-device Real-time Body Pose tracking. 2020.
[2] Di Paolo, Stefano, et al. Rehabilitation and return to sport assessment after anterior cruciate ligament injury: quantifying joint kinematics during complex high-speed tasks through wearable sensors. Sensors, 2021, 21. Jg., Nr. 7, S. 2331.
[3] Jones, Bruce H.; COWAN, David N.; KNAPIK, Joseph J. Exercise, training and injuries. Sports Medicine, 1994, 18. Jg., S. 202-214.
[4] Mehrizi R, Peng X, Metaxas DN, et al. Predicting 3-D Lower Back Joint Load in Lifting: A Deep Pose Estimation Approach. IEEE Trans Hum-Mach Syst 2019;49(1):85–94
[5] Paulich, Monique, et al. Xsens MTw Awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3D kinematic applications. Xsens: Enschede, The Netherlands, 2018.
[6] Trinidad-Fernández M, Cuesta-Vargas A, Vaes P, et al. Human motion capture for movement limitation analysis using an RGB-D camera in spondyloarthritis: a validation study. Med Biol Eng Comput 2021;59(10):2127–2137.

Fachgebiet Künstlich Intelligente Systeme der Medizin * AI Systems in Medicine Lab
www.kismed.tu-darmstadt.de * Prof. Dr.-Ing. Christoph Hoog Antink * hoogantink@kismed.tu-darmstadt.de

KIS*MED