A Data-Based Parametric Biomechanical Running Model driven by Pose Estimation

- A proof of concept study -

Master´s Thesis

Jesper Bang Hansen
Nikolaj Hoelgaard Kristiansen

Aalborg University
Sports Technology

Department of Health Science and Technology
Aalborg University
https://www.hst.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
A Data-Based Parametric Biomechanical Running Model driven by Pose Estimation - A proof of concept study

Project Period:
Spring Semester 2022

Project Group:
hst-22-idrtek-4-06

Authors:
Jesper Bang Hansen
Nikolaj Hoelgaard Kristiansen

Supervisors:
John Rasmussen
Mark De Zee

External Supervisors:
Christian Ingwersen (TrackMan)
Mikkel Jul Hansen (TrackMan)

Date of Completion:
June 1, 2022

Abstract:

This study investigated the concept of using a statistical parametric running model driven by knee and hip flexion from single-view pose estimation to recreate running patterns comparable to a marker-based motion capture reference. 51 treadmill running trials from 18 male subjects were recorded with a smartphone and marker-based motion capture simultaneously. Joint angles extracted from pose estimation were transformed to Fourier coefficients, where coefficient a_1 and b_1 were used as constrained parameters in the parametric running model along with subject height, weight, gender, age, angular stride frequency and running speed. The comparison showed excellent Pearson's correlation coefficients (r $>$ 0.90) for hip, knee and plantar flexion and vertical center of mass position and strong correlations (0.67 $<$ r $\leq$ 0.90) for hip abduction and elbow flexion. Root mean square difference (RMSD) for knee and plantar flexion and hip abduction were between 4.0^o and 8.5^o, hip and elbow flexion RMSDs were between 10.0^o and 14.9^o, while vertical center of mass RMSD was 0.021m. In conclusion, the investigated concept was capable of estimating realistic running kinematics comparable to the marker-based reference. However, it would require an increased number of accurately measured input parameters to reduce the magnitude errors of the estimated kinematic model outputs.

A Data-Based Parametric Biomechanical Running Model driven by Pose Estimation - A proof of concept study

Jesper Bang Hansen[a], Nikolaj Hoelgaard Kristiansen[a]

[a]*Department of Health Science and Technology, Aalborg University, Fredrik Bajers vej 7 D2, Aalborg, 9000, Denmark*

Abstract

This study investigated the concept of using a statistical parametric running model driven by knee and hip flexion from single-view pose estimation to recreate running patterns comparable to a marker-based motion capture reference. 51 treadmill running trials from 18 male subjects were recorded with a smartphone and marker-based motion capture simultaneously. Joint angles extracted from pose estimation were transformed to Fourier coefficients, where coefficient a_1 and b_1 were used as constrained parameters in the parametric running model along with subject height, weight, gender, age, angular stride frequency and running speed. The comparison showed excellent Pearson's correlation coefficients (r > 0.90) for hip, knee and plantar flexion and vertical center of mass position and strong correlations (0.67 < r ≤ 0.90) for hip abduction and elbow flexion. Root mean square difference (RMSD) for knee and plantar flexion and hip abduction were between 4.0^o and 8.5^o, hip and elbow flexion RMSDs were between 10.0^o and 14.9^o, while vertical center of mass RMSD was 0.021m. In conclusion, the investigated concept was capable of estimating realistic running kinematics comparable to the marker-based reference. However, it would require an increased number of accurately measured input parameters to reduce the magnitude errors of the estimated kinematic model outputs.

Keywords: Motion analysis, Markerless motion capture, Running, Kinematics

1. Introduction

Movement analysis in sports and health science has developed considerably in recent years, regarding both performance optimization and injury prevention. This development is largely due to improved technology and computational power [4]. When investigating the kinematics of human movement, marker-based optical motion capture remains the highest measurement standard [11]. Despite the high accuracy of this method, its complex setup and dependency on laboratory settings limits the availability of movement analysis in a natural environment. Furthermore, marker placement can be prone to positioning variability from the operator [5]. Reflective markers on the body may move unintentionally due to soft tissue artifacts, they may fall off due to fast accelerations or sweat, while they also may hinder the natural movement. The need for in-field kinematic analysis, has sparked the development of other motion capture solutions based on IMUs. However, these solutions can be prone to signal drift, while also hindering natural movement, due to IMUs being placed on several body segments [1]. Recently, deep learning-based markerless optical motion capture systems have also been suggested as an alternative to the traditional laboratory reference, showing similar joint angle patterns, when compared to the marker-based solutions in walking trials [9]. However, these systems often require a setup of several specific cameras, and are yet to be evaluated in more dynamic and sports-related movements.

Another improving non-laboratory optical markerless motion capture method is pose estimation based on monocular image input. This is a computer vision technique that detects and tracks the articulated human joint locations from simple image or video input [18]. Pose estimation with monocular image input has shown promising results in estimating body landmarks, e.g. head, shoulder, elbow, hip, knee and ankle [19]. Furthermore, studies have shown promising results in joint angle estimation based on the predicted landmarks [3, 12]. However, this method alone remains insufficient for full-body kinematic analysis, since not all degrees of freedom can be covered with high accuracy [6]. Thus, when wanting to develop a simple in field movement analysis method, it is relevant to investigate solutions capable of generating high quality full-body kinematics based on minimal and possibly noisy input data.

Rasmussen et al. have developed a parametric running model, that based on a large database of running trials can predict the most probable running pattern from minimal subject and trial-specific input [13]. Based on kinematic and anthropometric data from 285 running trials, the model is subject to Principal Component Analysis, which enables variation of the parameters to recreate plausible running patterns. The model can be considered as a conditional likelihood algorithm, where each kinematic running parameter is represented with five Fourier coefficient pairs. Representing running kinematics as Fourier series has shown to be valid and is more compact than a temporal representation [15]. This allows the model to handle large amounts of data with a low computational cost. Thus, the aim of the current study was to investigate the concept of using a statistical parametric running model driven by knee and hip flexion from single-view pose estimation to recreate a running pattern comparable to a marker-based motion capture reference.

2. Methods

The methodological structure of the current study followed the steps shown in Fig 1, with the purpose of covering the performance of a parametric running model driven by hip and knee joint angles estimated through pose estimation (Model-PE) by comparing it to the results of a parametric running model driven by hip and knee joint angles derived from marker-based motion capture (Model-MB) and the results from a golden standard reference of full-body marker-based motion capture measurement (Reference).

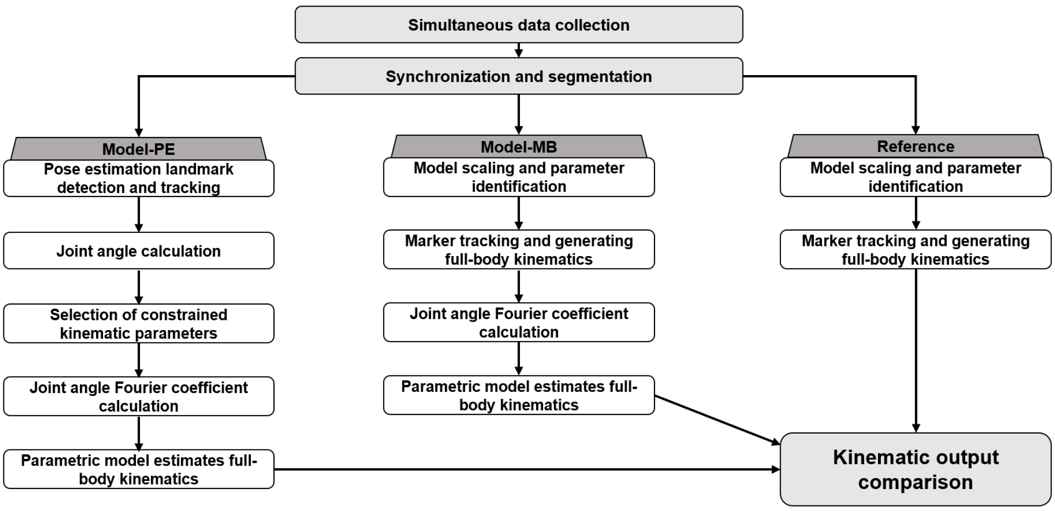

Figure 1: A flow diagram showing the step-by-step workflow of the data collection and processing used in the three models; Model-PE, Model-MB and Reference.

2.1. Data collection

A total of 18 male subjects ran three trials each on a treadmill. Each subject was equipped with 40 reflective markers (see Appendix A.5) and for each trial, marker trajectories were recorded by a 3D motion capture setup with 8 cameras (Qualisys Oqus M3, Qualisys AB, Gothenburg, Sweden) at 240 frames per second. Each trial was simultaneously recorded by a regular smartphone camera (iPhone SE, Apple, California, USA) at 240 frames per second. The smartphone camera was placed 2.30 meters from the treadmill fixed by a tripod at a height of 1.44 meters, with the camera view in the sagittal plane of the subject, with the hip of the subject as the viewing center. A clapping board with reflective markers was used in the viewing field of both camera systems to allow for synchronization. The subjects chose three different running speeds (7 - 15.5 km/h), they felt comfortable running in. In each trial, subjects ran on the treadmill until steady state was attained, and then a period of 10 s was recorded. Three trials were excluded due to poor video quality, hence 51 running trials were included for analysis.

2.2. *Segmentation*

The recorded trials were trimmed to begin at the first right foot heel strike and end at the second right foot heel strike. In this way, data corresponded to the same stride cycles as in the parametric running model [13]. Heel strikes were identified using the peak acceleration of the heel marker. Hereafter, the corresponding frames were extracted from the video files.

2.3. *Marker-based kinematics*

The marker trajectories were imported as c3d files to the AnyBody Modeling System (version 7.3), filtered with a low pass second order Butterworth filter of 10 Hz and used to drive a full body musculoskeletal model (AMMR.v.2.3.0). The musculoskeletal model comprised 67 anatomical segments, 52 anatomical joints and 104 joint degrees-of-freedom. Utilizing the parameter identification algorithm described by Andersen et al. [2], the time-series representation of the full-body kinematics were estimated.

2.4. *Pose estimation and joint angle calculation*

Data processing of the video files was done in Python v 3.8.8. First, the trimmed video files were imported using OpenCV v. 4.5.5.62. The video files were stored as matrices containing RGB information of the pixels in each frame. Then the pose estimation was performed on each frame of the video files, using the MediaPipe Pose Python package. As the input was a video stream, the static image mode was set to false, meaning that person detection was only run on the first frame. Hereafter, the algorithm tracked each landmark through the rest of the frames and only ran another detection if landmarks were lost. Minimum detection and tracking confidence were both set to 0.9. Model complexity was set to 2, since it has the highest accuracy and computation time was not a concern since no live processing was done.

The hip and knee joint flexion angles θ_j were calculated in 2D by subtracting the two segment angles θ_1 and θ_2 combining in the specific joint:

$$\theta_j = \theta_1 - \theta_2$$

Here each segment angle is calculated in respect to the horizontal axis using the arcus tangent function:

$$\theta = arctan(\frac{y_2 - y_1}{x_2 - x_1})$$

where y_1 and x_1 are the 2D properties of the central joint of interest, in example the knee joint. While y_2 and x_2 are the 2D properties of either the hip or ankle joint, depending on whether the shank or thigh segment angle is being calculated.

When the time-series of the hip and knee flexion angles were recovered, they were subject to Fourier analysis. Here each joint angle curve was approximated by a sum of sine and cosine terms:

$$y(t) = a_0 + \sum_{i=1}^{5} a_i cos(i\omega t) + \sum_{i=1}^{5} b_i sin(i\omega t)$$

where ω is the angular stride frequency, 5 is the number of coefficient pairs. The Fourier coefficients a_0, a_i and b_i were found using a Riemann integral over the period of the function following the method of Skejø et al. [15].

2.5. *Parametric biomechanical running model*

The parametric biomechanical running model of Rasmussen et. al. contains 104 joint degrees of freedom, each represented by five Fourier coefficient pairs [15]. Furthermore, it consists of subject-specific anthropometric data and trial-specific metadata, i.e., gender and running speed. In total, this gives 1224 primal parameters describing the running pattern [13].

The Fourier coefficients of the hip and knee flexion angles were used as constrained primal parameters in the parametric running model along with the angular stride frequency, anthropometric data of the total body height and body mass and metadata of running speed, gender and age. The constrained primal parameters were defined in the Python program and the remaining degrees of freedom to recreate the most probable running pattern were estimated by solving the optimization problem described in [13].

2.6. *Selection of kinematic input*

Before evaluating the performance of the pose estimation driven running model, the number of constrained Fourier coefficient pairs were determined by evaluating the Pearson's correlation coefficient (r) and Root Mean Squared Difference (RMSD) between the predicted knee and hip flexion curves and the marker-based reference. Fourier coefficients were evaluated in pairs with the same coefficient number (a_1 and b_1 = pair 1 etc.). In relation to the coefficient pair selection, the number of constrained coefficient pairs were only increased, if the higher order of pairs resulted in a higher correlation coefficient and lower RMSD without increasing the range.

2.7. *Comparison*

When constraining the parametric running model with input derived from pose estimation, the hip and knee flexion curves would not be identical to the ground truth. Thus, a model driven by a_0 and 5 constrained coefficient pairs derived from the marker-based reference was created to make a model driven by ground truth hip and knee flexion Fourier coefficients. This gives 2 models; Model-PE constrained with coefficients derived from pose estimation and Model-MB constrained with a_0 and five coefficient pairs, derived from the marker-based reference. Both models used the same anthropometric and meta data input and were compared to the full-body musculoskeletal reference model driven by marker-based motion capture, see Fig 1.

Plantar flexion, elbow flexion, hip abduction angles and vertical center of mass position (vCoM) were selected to compare the running pattern predicted by Model-PE and Model-MB with the marker-based reference. The Fourier coefficients of the selected parameters were transformed back to the time series domain, where Pearson's correlation coefficient, RMSD and Mean Average Error (MAE) were used to compare the shape and magnitude difference between the two models and the marker-based reference. The Pearson correlation coefficients were categorized according to Taylor as "weak" ($r \leq 0.35$) , "moderate" ($0.35 < r \leq 0.67$), "strong" ($0.67 < r \leq 0.90$) and "excellent" ($r > 0.90$) [17].

3. Results

3.1. *Kinematic model input*

Table 1: The median (min - max) of the Pearson's correlation coefficients and root mean squared differences between the pose estimation driven joint angle predictions of the right and left knee and hip flexion and the measured reference. 1 - 5 Fourier coefficient pairs were constrained. 1 pair = a_1 and b_1, 2 pairs = a_1, a_2, b_1 and b_2 etc.

	Degree of freedom							
	Right knee flexion		Left knee flexion		Right hip flexion		Left hip flexion	
Pairs	r	RMSD	r	RMSD	r	RMSD	r	RMSD
1	0.979 (0.915 - 0.997)	8.5 (2.8 - 18.0)	0.989 (0.935 - 0.997)	7.4 (2.9 - 11.6)	0.984 (0.944 - 0.997)	11.7 (4.4 - 20.8)	0.981 (0.913 - 0.997)	12.3 (6.7 - 21.3)
2	0.962 (0.874 - 0.993)	10.3 (4.7 - 18.6)	0.981 (0.903 - 0.996)	9.4 (6.9 - 12.7)	0.985 (0.909 - 0.998)	9.7 (3.5 - 18.8)	0.974 (0.880 - 0.997)	12.1 (3.6 - 18.6)
3	0.965 (0.881 - 0.991)	12.8 (5.7 - 26.3)	0.976 (0.916 - 0.993)	11.1 (7.7 - 23.6)	0.983 (0.912 - 0.996)	11.4 (3.7 - 28.8)	0.965 (0.880 - 0.988)	12.9 (2.6 - 23.7)
4	0.965 (0.880 - 0.990)	10.9 (5.4 - 26.8)	0.979 (0.917 - 0.991)	10.4 (6.5 - 22.3)	0.983 (0.910 - 0.997)	8.6 (2.0 - 21.6)	0.970 (0.880 - 0.990)	11.5 (3.4 - 22.7)
5	0.965 (0.880 - 0.990)	11.0 (4.8 - 25.7)	0.979 (0.917 - 0.990)	9.8 (5.4 - 32.1)	0.982 (0.912 - 0.996)	7.7 (2.3 - 26.7)	0.969 (0.881 - 0.989)	10.6 (5.9 - 27.5)

Constraining a_0 resulted in lower correlation coefficients and higher magnitude errors in all combinations and was therefore not considered for Model-PE. Table 1 shows that the highest correlation and lowest RMSD was found when only constraining a_1 and b_1 for both knee flexion angles. For the hip flexion angles, the highest correlations were found when constraining one coefficient pair for the left hip and two coefficient pairs for the right hip. The lowest RMSDs were found when constraining all five coefficient pairs. Thus, Model-PE was driven by constrained a_1 and b_1 coefficients. Hereby, the unconstrained coefficient pairs 2 to 5 became part of the running model estimation, resulting in the time series joint angle curves shown in Fig 2.

3.2. Model performance

Figure 2: Left and right knee and hip flexion means from Model-PE and the marker-based reference with Pearson's correlation coefficient, RMSD and MAE median. The horizontal axis is normalized to 100% gait cycle and the vertical axis is joint angle degrees.

Excellent correlations were found for both knee and hip flexion on both sides. However, the hip flexion angle was generally estimated smaller for Model-PE than the measured reference, with a high RMSD compared to the range-of-motion. As shown in Fig 3 both Model-PE and Model-MB plantar flexion angles showed excellent correlations for both sides with slightly higher correlations for Model-MB. Both models showed similar RMSD values, with a marginally smaller magnitude error for the left ankle. For the elbow flexion both models showed strong correlations for both sides, with model-PE having the highest correlation coefficient on the left side and Model-MB on the right side. On the left side Model-PE showed the lowest magnitude error, while Model-MB had a slightly lower RMSD on the right side. For the hip abduction both models showed strong correlations, with Model-MB having the highest correlation coefficients for both sides. However, Model-PE showed the lowest magnitude error for both sides.
Both Model-PE and Model-MB showed excellent correlations for the vCoM, with the highest correlation coefficient for Model-MB. The models showed similar magnitude errors, see Fig 4.

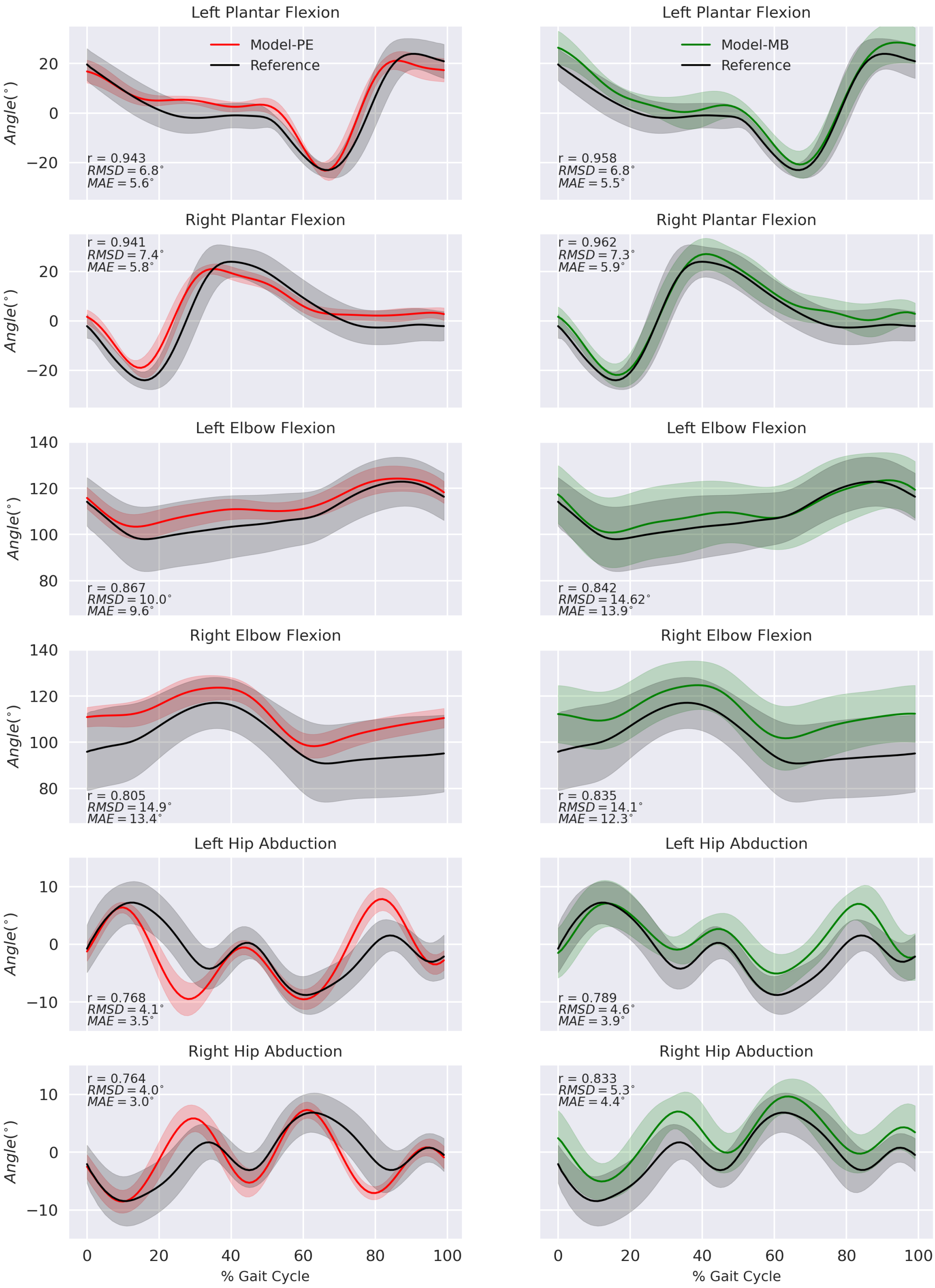

Figure 3: Left and right plantar flexion, elbow flexion and hip abduction means and standard deviations from Model-PE, Model-MB and marker-based reference with Pearson's correlation coefficient, RMSD and MAE median. The horizontal axis is normalized to 100% gait cycle, and the vertical axis is degrees.

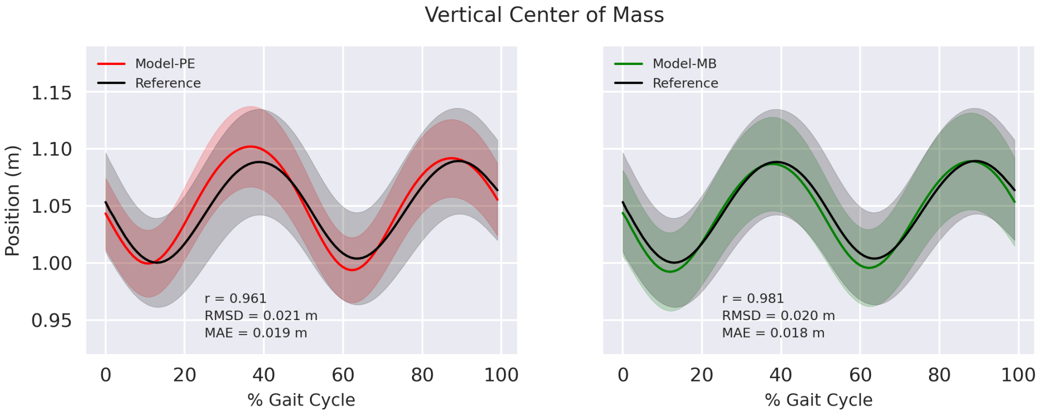

Figure 4: Vertical Center of mass means and standard deviations from Model-PE, Model-MB and marker-based reference with Pearson's correlation coefficient, RMSD and MAE median. The horizontal axis is normalized to 100% gait cycle, and the vertical axis is in meters.

4. Discussion

The aim of this study was to investigate the concept of using a statistical parametric running model driven by knee and hip flexion angles from pose estimation to recreate a running pattern comparable to a reference from marker-based motion capture. The results of Model-PE and Model-MB showed strong to excellent correlations compared to the marker-based reference on all evaluation parameters. Knee and hip flexion changes significantly based on the running velocity and stride frequency [14, 8]. Thus, these DOFs hold much information and are of high importance, when estimating the running pattern based on minimal model input data. Regarding the selection of the number of constrained coefficient pairs in Model-PE, using two coefficient pairs showed the highest median correlation for right hip flexion, while five coefficient pairs showed the lowest median RMSD for hip flexion for both sides. However, the range of the right hip flexion with two coefficient pairs showed lowest minimum correlation (r = 0.909) of all amounts of pairs, while the maximum RMSDs for the right and left hip flexion was higher with five constrained coefficient pairs (right hip flexion: RMSD = 26.7^o, left hip flexion: RMSD = 27.5^o) compared to using one pair (right hip flexion: RMSD = 20.8^o, left hip flexion: RMSD = 21.3^o). It was a priority not to increase the number of constrained coefficient pairs, if this increase would not influence both the Pearson's correlation, RMSD and range positively, to ensure that an eventual increase would equally improve the model performance. Hence, constraining one coefficient pair for the hip and knee flexion was chosen for Model-PE. Constraining only one coefficient pair results in a more general running pattern estimation, thus decreasing the trial specificity of the model output. This influence is shown by the standard deviations of each joint angle output of Model-PE which were smaller compared to Model-MB, see Fig 2 and 3. Hence, when increasing the amount of constrained Fourier coefficients in the model, the specificity of the output will also increase. Yet, doing this would equally require a high accuracy of the pose estimated joint angles, to ensure that constraining more coefficients would not result in inaccurate running pattern estimations.

vCoM can be considered as a product of multiple kinematic parameters in the lower body, e.g. plantar, knee and hip flexion angles. vCoM is well predicted by the two models with excellent correlations (Model-PE: p = 0.961, Model-MB: p = 0.981) and very low magnitude errors (Model-PE: RMSD = 0.021m, Model-MB: RMSD = 0.020m). This extends the possibility to investigate the vertical ground reaction forces that are highly related to the vertical accelerations of the CoM and lower body kinematics [7]. Thus, allowing future investigations of muscle- and joint reaction forces to help explain causes of pain and overuse injuries by extensive load at specific joints or tissue.

Stenum et al. have proposed a single-view pose estimation setup to investigate kinematic gait parameters in slow walking. They showed knee and hip correlations similar to the current study, while the plantar flexion in the current study showed higher correlations [16]. This indicates that the concept of using Model-PE to estimate lower body kinematics performs better than direct joint angle estimation using a single camera view. The knee and hip flexion MAE in [16] showed lower values compared to the current study. However, this can be explained by the higher limb movement velocity and larger range-of-motion in the running gait compared to slow walking. Kanko et al investigated walking gait kinematics using the multi-view markerless Theia3D setup, showing plantar and hip flexion RMSDs similar to the current study, while showing lower RMSDs for knee flexion and hip abduction [10]. A multi-view setup enables direct assessment of joint angles in the frontal plane that possibly enables a better hip abduction estimation compared to Model-PE, which is only kinematically constrained with parameters from the sagittal plane. Model-PE showed lower correlations and higher MAE values for plantar, knee and elbow flexion compared to Pagnon et al. who evaluated their multi-view pose estimation model Pose2Sim in running [11]. However, Pose2Sim was based on 8 cameras using a textured mesh of each participant created from 68 cameras. Thus, being restricted to a complex setup compared to the single view pose estimation setup used in the current study. For the hip flexion and abduction Model-PE showed higher correlations and lower MAE compared to Pose2Sim. Despite using a more simplistic setup, the Model-PE performed better on specific parameters compared to a complex multi-view pose estimation setup.

Model-MB showed lower correlations for hip abduction (left: r = 0.789, right: r = 0.833) and elbow flexion (left: r = 0.842, right: r = 0.835) compared to plantar flexion (left: r = 0.958, right: r = 0.962) and vCoM (r = 0.981). Thus, a more accurate estimation of these DOFs would require an increase in the amount of constrained parameters in the parametric running model. The relationship between the combination of constrained parameters and the output of the parametric running model is unknown, since the model estimation holds a lot of complexity, where different parameters are statistically correlated, thus influencing each other in the internal algorithms of the statistical model. However, it seems feasible to constrain DOFs of the upper-body, when wanting to increase the accuracy and trial specificity of upper-body parameters.

In conclusion, this study has shown that a statistical parametric running model driven by knee and hip flexion angles derived from single-view pose estimation is capable of recreating running kinematics comparable to a marker-based reference. The influence of adding accurate upper-body DOFs from pose estimation to the constrained model input should be investigated to improve overall model performance.

Appendix A. Sample Appendix Section

Marker	Position
RFHD	Right forehead
LFHD	Left forehead
RBHD	Right back head
LBHD	Left back head
C7	7th Cervical Vertebrae
RSHO	R. Acromio-clavicular joint
LSHO	L. Acromio-clavicular joint
CLAV	Jugular notch
RUPA	Right Triceps brachii
LUPA	Left Triceps brachii
RELB	Right elbow lateral epicondyle
LELB	Left elbow lateral epicondyle
RFRA	Right Radius
LFRA	Left Radius
RWRB	Right wrist bar finger side
RWRA	Right wrist bar thump side
RFIN	Right top of hand sensor
LWRB	Left wrist bar finger side
LWRA	Left wrist bar thump side
LFIN	Left top of hand sensor
STRN	Xiphoid process of sternum
T10	10th thoracic vertebrae
RASI	R. anterior superior iliac spine
LASI	L. anterior superior iliac spine
RPSI	R. posterior superior iliac spine
LPSI	L. posterior superior iliac spine
RTHI	Right thigh
LTHI	Left thigh
RKNE	Right knee lateral epicondyle
LKNE	Left knee lateral epicondyle
RTIB	Right tibia
LTIB	Left tibia
RHEE	Right calcaneus
RANK	Right lateral malleolus
RMT5	Right 5th metatarsal
RTOE	Right metatarsus
LHEE	Left calcaneus
LANK	Left lateral malleolus
LMT5	Left 5th metatarsal
LTOE	Left metatarsus

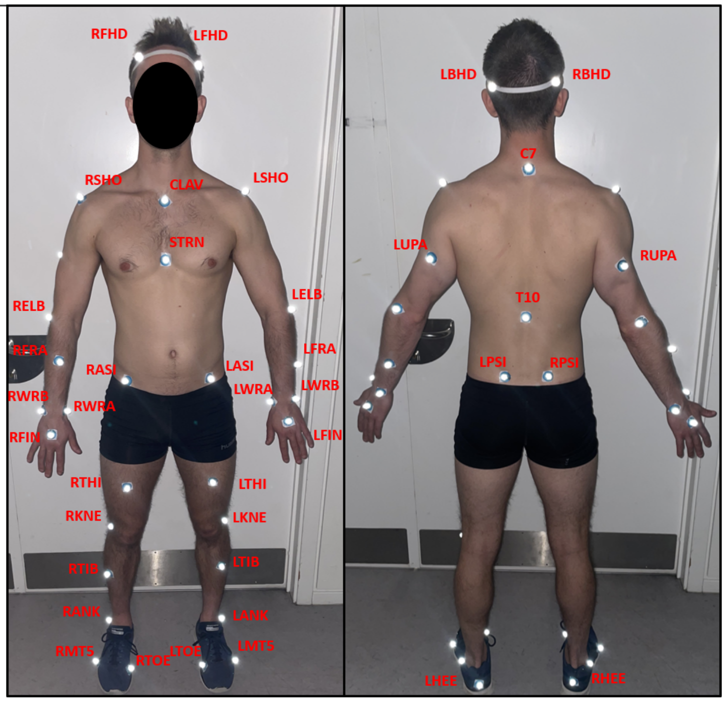

Figure A.5: Marker placement protocol for the marker-based motion capture reference.

References

[1] Ahmad, N., Ghazilla, R.A.R., Khairi, N.M., Kasi, V., 2013. Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications. International Journal of Signal Processing Systems , 256–262URL: http://www.ijsps.com/index.php?m=contentc=indexa=showcatid=32id=65, doi:10.12720/ijsps.1.2.256-262.

[2] Andersen, M.S., Damsgaard, M., MacWilliams, B., Rasmussen, J., 2010. A computationally efficient optimisation-based method for parameter identification of kinematically determinate and over-determinate biomechanical systems. Computer methods in biomechanics and biomedical engineering 13, 171–183.

[3] Cheng, Y., Yang, B., Wang, B., Tan, R.T., 2020. 3D Human Pose Estimation using Spatio-Temporal Networks with Explicit Occlusion Training URL: https://arxiv.org/abs/2004.11822, doi:10.48550/ARXIV.2004.11822. publisher: arXiv Version Number: 1.

[4] Cust, E.E., Sweeting, A.J., Ball, K., Robertson, S., 2019. Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. Journal of sports sciences 37, 568–600.

[5] Della Croce, U., Cappozzo, A., Kerrigan, D.C., 1999. Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles. Medical & Biological Engineering & Computing 37, 155–161. URL: http://link.springer.com/10.1007/BF02513282, doi:10.1007/BF02513282.

[6] Desmarais, Y., Mottet, D., Slangen, P., Montesinos, P., 2021. A review of 3D human pose estimation algorithms for markerless motion capture. Computer Vision and Image Understanding 212, 103275. URL: https://linkinghub.elsevier.com/retrieve/pii/S1077314221001193, doi:10.1016/j.cviu.2021.103275.

[7] Fluit, R., Andersen, M., Kolk, S., Verdonschot, N., Koopman, H., 2014. Prediction of ground reaction forces and moments during various activities of daily living. Journal of Biomechanics 47, 2321–2329. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021929014002516, doi:10.1016/j.jbiomech.2014.04.030.

[8] Heiderscheit, B.C., Chumanov, E.S., Michalski, M.P., Wille, C.M., Ryan, M.B., 2011. Effects of Step Rate Manipulation on Joint Mechanics during Running. Medicine & Science in Sports & Exercise 43, 296–302. URL: https://journals.lww.com/00005768-201102000-00014, doi:10.1249/MSS.0b013e3181ebedf4.

[9] Kanko, R.M., Laende, E., Selbie, W.S., Deluzio, K.J., 2021a. Inter-session repeatability of markerless motion capture gait kinematics. Journal of Biomechanics 121, 110422. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021929021002025, doi:10.1016/j.jbiomech.2021.110422.

[10] Kanko, R.M., Laende, E.K., Davis, E.M., Selbie, W.S., Deluzio, K.J., 2021b. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. Journal of Biomechanics 127, 110665. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021929021004346, doi:10.1016/j.jbiomech.2021.110665.

[11] Pagnon, D., Domalain, M., Reveret, L., 2021. Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness. Sensors 21, 6530. URL: https://www.mdpi.com/1424-8220/21/19/6530, doi:10.3390/s21196530.

[12] Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M., 2019. 3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA. pp. 7745–7754. URL: https://ieeexplore.ieee.org/document/8954163/, doi:10.1109/CVPR.2019.00794.

[13] Rasmussen, J., Lund, M.E., Waagepetersen, R.P., 2020. Data-based parametric biomechanical models for cyclic motions, in: DHM 2020 - Proceedings of the 6th International Digital Human Modeling Symposium, IOS Press. pp. 372–379. doi:10.3233/ATDE200045.

[14] Schache, A.G., Blanch, P.D., Dorn, T.W., Brown, N.A.T., Rosemond, D., Pandy, M.G., 2011. Effect of Running Speed on Lower Limb Joint Kinetics. Medicine & Science in Sports & Exercise 43, 1260–1271. URL: https://journals.lww.com/00005768-201107000-00017, doi:10.1249/MSS.0b013e3182084929.

[15] Skejø, S.D., Lund, M.E., Stensvig, M., Kaae, N.M., Rasmussen, J., 2021. Running in circles: Describing running kinematics using Fourier series. Journal of Biomechanics 115, 110187. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021929020306114, doi:10.1016/j.jbiomech.2020.110187.

[16] Stenum, J., Rossi, C., Roemmich, R.T., 2021. Two-dimensional video-based analysis of human gait using pose estimation. PLOS Computational Biology 17, e1008935. URL: https://dx.plos.org/10.1371/journal.pcbi.1008935, doi:10.1371/journal.pcbi.1008935.

[17] Taylor, R., 1990. Interpretation of the correlation coefficient: a basic review. Journal of diagnostic medical sonography 6, 35–39.

[18] Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., Shao, L., 2021. Deep 3D human pose estimation: A review. Computer Vision and Image Understanding 210, 103225. URL: https://linkinghub.elsevier.com/retrieve/pii/S1077314221000692, doi:10.1016/j.cviu.2021.103225.

[19] Zhang, H., Ouyang, H., Liu, S., Qi, X., Shen, X., Yang, R., Jia, J., 2019. Human Pose Estimation with Spatial Contextual Information URL: https://arxiv.org/abs/1901.01760, doi:10.48550/ARXIV.1901.01760. publisher: arXiv Version Number: 1.

Worksheet

1. Introduction

The article of this study was prepared following the submission guidelines of the Journal of Biomechanics. This master's thesis was made in collaboration with the company TrackMan. TrackMan specializes in developing analysis solutions across different sports like golf, baseball, tennis, soccer and American football. However, golf is their primary business segment. Through their Doppler radar, they are capable of monitoring the launch of a golf ball. With each swing, the Doppler radar system measures different aspects from club movement to ball trajectory and speed. TrackMan are still yet to develop an easy usable solution for analyzing the athlete movement that produces the golf swing. There are solutions based on artificial intelligence available that are capable of analyzing human movement based on simple video recordings. These solutions are however still too inaccurate for valid full-body analysis of a golf swing.

Rasmussen et al. have recently developed a parametric running model based on a large set of marker-based motion capture data, capable of constructing a full body kinematic analysis based on minimal trial-specific input [1]. It is in the interest of TrackMan to evaluate whether or not it is possible to drive this model with minimal noisy data input from simple video recordings. If successful, it would be purposeful to replace the running model with a parametric golf model that could be integrated in the Trackman analysis system. Thus, the aim of this study was to investigate the concept of using a statistical parametric running model driven by joint angles from single-view pose estimation to recreate a running pattern comparable to a reference from marker- based motion capture.

The following worksheet will provide an elaborate description of the theory, rationale and process behind the investigation. Furthermore, it will include the choices and thoughts regarding the study background, methodological theory, data processing and considerations in regard to both the scientific work and the practical perspectives of TrackMan.

2. Pose estimation

Human pose estimation is a computer vision technique that predicts and tracks the articulated joint locations of a human body from an image or a sequence of images [2]. Essentially it is a way to capture a set of coordinates for the human joints and thereby recover the underlying kinematic structure of a person. The results of human 3D pose estimation can be 2D or 3D poses or 3D human models corresponding to an image or video [3]. In research of sports and health related biomechanics, marker-based motion capture systems are most often used to describe and analyze the kinematics of a subject. Despite the high accuracy of this method, its dependency on laboratory equipment and settings restricts the possibility of assessing subjects outside a laboratory. Furthermore, the need for placing small markers on the surface of a subject, would have an impact on the motion of the subject. And especially for analyzing kinematics in sports, these would influence the performance of the athlete. Thus, the method of reconstructing human kinematics from 3D pose estimation, could increase the ecological validity of movement analysis, and hopefully advance the possibility for analyzing performance in the field.

The main challenges concerning pose estimation include the variation of body poses, complicated backgrounds, diverse clothing appearance and occlusions. However, the reconstruction of human skeletal representation and joint coordinate estimates in 2D has made a lot of progress in recent years, due to the development of convolutional neural networks and accessible large-scale datasets [4, 5]. Because of this development, recent work has focused on further development of 3D pose estimation. Constructing a 3D human pose from 2D image sequences is appealing, due to the possible usability and the requirement of few resources. However, it suffers from the complexity that different 3D poses may correspond to the same 2D image because of the added depth [2]. The extraction of 3D poses from videos facilitates analysis of the performance of athletes and provides immediate feedback for their movement [6]. Thus, generating precise 3D pose estimations is advantageous, not only for single frame estimations, but even more for sequences of images. However, extracting poses from a sequence of images adds further challenges to the estimation problem. For example, background variation, occlusion, camera movements, fast motion, loose clothing, and change in light condition may cause pose shape and appearance of people to change over time [2].

Multi-view camera setups have been investigated as a solution for the depth ambiguity of 3D pose estimation. When using multi view, different camera angles are used to construct a human pose. Here, 3D joint coordinates can be derived using fusion between the cameras or triangulation of the 2D image coordinates in each view [7]. Different multi-view approaches have shown to slightly increase the accuracy of 3D pose estimation with an average joint position error of 20-30 mm, while a monocular view, where only one camera angle is used for input, has reached about 40 mm [8]. Meaning, that despite access to epipolar geometry and 3D information, even the best multi-view setups present results comparable to those of the best monocular methods. Thus, the most discussed issue concerning human 3D pose estimation remains, that the accuracy seems insufficient for applications like motion analysis, which are essential for biomechanics and sports performance analysis [8]. The uncertainty that is still related to 3D pose estimation verifies looking for other measurement approaches that are valid and comparable to marker-based motion capture systems.

2.1. 3D Pose estimation using MediaPipe

In the current study, the parametric running model received kinematic input derived from the open source 3D Pose detection python package MediaPipe Pose. MediaPipe Pose is a machine learning solution for single person 3D pose estimation developed by Google, customizable to use in RGB images, video sequences and live recordings. The package utilizes a two-step detector-tracker machine learning pipeline, where it firstly detects the pose region-of-interest within the frame. The pose detection model initially predicts two keypoints, one describing the human body center and one placed above the subject's head, describing the radius of a circle circumscribing the subject. This circle defines the region-of-interest, where pose landmarks can be identified, see Fig 1.

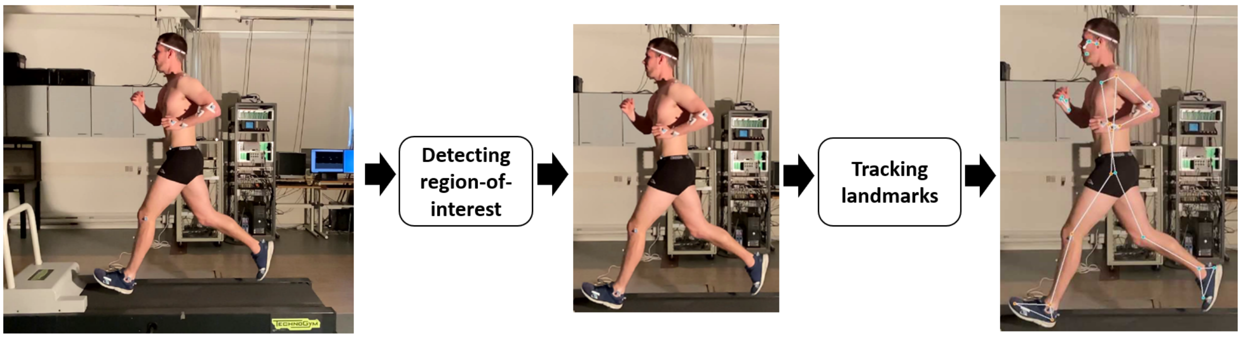

Figure 1: The two-step detector-tracker pose estimation pipeline.

After subject detection, a tracking model predicts 33 3D pose and face landmarks within the region-of-interest, these are shown in Fig 2. In the subsequent frames, the tracker predicts the pose and face landmarks based on both the image input and the previous frames, meaning that the detection model only is activated if needed. This would be, when the tracker is not capable of identifying body pose presence in the previous frame.

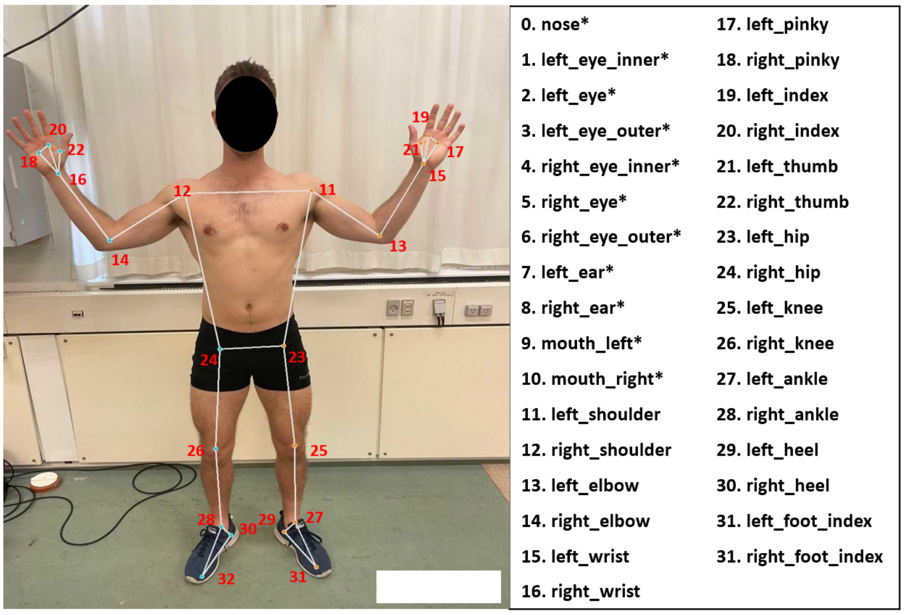

Figure 2: Example of the positioning and names of body landmarks in MediaPipe Pose. Landmark positions 0-10 marked by (*) are deliberately hidden in the image to ensure subject privacy.

2.2. Adjustable MediaPipe code arguments

To initialize the pose detection model in a python environment we call the setup function mp.solutions.pose.Pose(), where a few different input arguments can be adjusted to fit the specific requirements and purpose of the pose estimation. In the current study all input arguments were adjusted with the purpose of getting the most accurate pose estimation, without concerning processing time [9]:

- **static_image_mode**: A boolean value. When set to false, the detector is only invoked if needed. If set to True, the detector is activated in every input frame. Hence this value should be set to True in a case of working with a file containing unrelated images in each frame. In the current project it was set to the default value of False, since working with video sequences.
- **min_detection_confidence**: A detection model confidence interval ranging from 0.0 - 1.0 with the default value of 0.5. In the current study this was set to 0.9, meaning that the detector needed a prediction confidence of greater or equal to 90% to be considered a positive prediction.
- **min_tracking_confidence**: A tracking model confidence interval ranging from 0.0 - 1.0 with the default value of 0.5. In the current study, this was set to 0.9 meaning that if the confidence of the tracked landmark is greater or equal to 90% then the tracked landmark position is valid. If the confidence is below 90%, then the detector is invoked in the next frame. Increasing this value increases the robustness of the pose estimation, but also the latency.
- **model_complexity**: MediaPipe is capable of using three different pose landmark models, with different complexity (0, 1 and 2) with value 1 as default. The higher the value, the more complex a model will be used, and the more accurate results will be. However this will equally increase the latency. In the current study, the value of 2 was chosen to increase the accuracy of the model's pose estimation.
- **smooth_landmarks**: A boolean value, with default value True when static_image_mode set to False. In the current study set to True, meaning that pose landmarks across different frames are filtered to reduce noise.

3. Describing running kinematics with Fourier series

In most biomechanical research, running kinematics are represented as joint angle curves in the time-series domain. This allows for visual interpretation using graphs, often with the time normalized to 100% of the gait cycle. However, it is difficult to reduce the time-series data to a few simple numbers to analyze the whole cycle. Therefore, kinematic data is often compared in qualitative terms or quantitatively at specific points in the gait cycle. Furthermore, it is troublesome to archive large amounts of kinematic data represented in the time-domain. The Fourier transform instead allows kinematic data of cyclic motions to be represented as a sum of sine and cosine waves:

$$y(t) = a_0 + \sum_{i=1}^{n} a_i cos(i\omega t) + \sum_{i=1}^{n} b_i sin(i\omega t)$$

Where ω is the angular stride frequency, n is the number of coefficient pairs and a and b are the Fourier coefficients. The value of the Fourier coefficients are given by:

$$a_0 = \frac{1}{T}\int f(t)dt \qquad a_i = \frac{2}{T}\int_0^T f(t)cos(\omega it)dt \qquad b_i = \frac{2}{T}\int_0^T f(t)sin(\omega it)dt$$

Where T is the stride time, f(t) is the joint angle at time t and i is the coefficient pair number [10]. As the order of the coefficient pairs increases, the sine and cosine curves are multiplied with the coefficient order, i.e. a_2 is multiplied with a cosine wave of two times the angular stride frequency, ω, and b_3 is multiplied with a sine wave of three times ω. Hence, the lower orders (e.g. n=1 or n=2) describe the general shape of the curve, whereas the higher orders of the coefficients describe more detailed fluctuations of the joint angle data [11]. Skejø et al. showed that joint angles from running trials can be represented with the use of 5 Fourier coefficient pairs. This allows for a very compact representation of running kinematics, which makes it possible to archive a large amount of data in a small size file.

4. Data-Based Parametric Biomechanical Models driven by 3D pose estimation

The 3D pose estimation methods available today are still too inaccurate for applicable motion analysis comparable to the standard of marker-based motion capture systems. Rasmussen et al. have developed a data-based parametric biomechanical running model, which utilizes that running kinematics can be described by Fourier series [1, 10].

When analyzing skilled and repetitive movements like running, swimming, or cycling, there seems to be a high degree of intersubjective kinematic similarity that can be described from connections between movement patterns and anthropometry. For instance, a positive correlation between leg length and step length seems plausible. Knowing these connections, it is potentially possible to reduce the need for subject-specific model input. Based on this context, Rasmussen et al. developed a unified statistical model for anthropometry and movements for cyclic motions. The statistical model was made based on 285 treadmill running trials (180 from male and 105 from female subjects) collected from 80 subjects using an optical motion capture system. Each trial was converted into a full-body musculoskeletal model that comprises 67 anatomical segments, 52 anatomical joints and 104 joint degrees-of-freedom. Over a typical running cycle, the variation of each joint degree of freedom was recovered from the model, and the time series functions were parameterized with Fourier series restricted by 11 Fourier coefficients. Meaning that each running trial is represented by 11 coefficients per degree-of-freedom, subject-specific segment lengths and meta data such as gender, age, weight, height etc. which in total comprises 1224 parameters. All 1224 parameters are collected in a large matrix, X, where each row represents a trial and each column is a parameter, x. The running parameters are statistically correlated, thus it is not possible to use the matrix X directly as a parametric running model. Specifying single running parameters, x, would simply lead to unrealistic running patterns. Therefore, the matrix, X, is subjected to principal component analysis (PCA):

$$y = A(x - \mu)$$

Here, the PCA transformation matrix A is used to transform a set of running parameters, x, and a vector of mean values from X, μ, into principal components, y. The transformed parameters, y, are uncorrelated and can therefore be changed within a statistical reasonable range. The PCA transformation matrix, A, has orthonormal properties, which allows an opposite transformation from transformed parameters, y, back to primal space, x, where the parameters have a physical interpretation i.e. running speed or Fourier coefficients of knee flexion angle.

$$x = A^T y + \mu$$

The PCA of the matrix, X, leads to a parametric running model, where the transformed parameters, y, can be varied and then transformed back to running parameters, x, to create new running patterns. To utilize this, Rasmussen et al. formulated an optimization problem, where they found that the most probable running pattern is found when y is minimized with constraints of some of the running patterns, x:

Minimize $\qquad y^T C{-}1y$

Subject to $\qquad x_j = f_j for j \in J$

Here x_j can be expressed in terms of y, which gives a quadratic optimization problem with the transformed variables, y, as linear constraints:

Minimize $\qquad y^T C{-}1y$

Subject to $\qquad \Sigma_i a_{ij} y_i + \mu_j = f_j for j \in J$

In the context of this project, this makes it possible to constrain the parametric running model with running parameters, x, that are Fourier coefficients of joint angles extracted from pose estimation, measured segment lengths and meta info i.e. gender or running speed, and then predict the remaining running parameters to make the most probable running pattern.

5. Protocol considerations

In order to standardize the experimental protocol, different setups were investigated and compared with the purpose of creating MediaPipe joint angle estimates as similar as possible to the reference. Thus, the many pose estimation challenges were considered and minimized through the setup.

5.1. Camera position and setup

When adjusting the experimental protocol and setup, a relevant consideration was the camera view of the subject. Different camera view positions were considered and tested through the pilot trials; frontal, back, sagittal, obliquely from the front and back. Results clearly showed most accurate relevant joint positions and angles from a view obliquely from the front and directly from a sagittal view. The oblique view had the advantage of getting a high pose landmark detection confidence on all 33 landmarks. While the sagittal view was less confident in regard to detecting the landmarks on the body on the opposite side of the camera. For example, when recording from the left side of the subject, the right elbow or shoulder would often be hidden behind the trunk. Thus, landmarks on the right side of the body would often be predicted on behalf of the left side. However, an advantage of the sagittal view was that it was less complicated to make the global coordinate system of the pose estimation agree with the local movement plane of the subject. Furthermore, with an oblique view, the accuracy of the pose estimation was very dependent on a valid prediction of all 3 dimensions, including the depth. However, a sagittal view showed great results for the knee and hip flexion, since these could be estimated based on the 2D view in this plane, excluding the importance of the depth estimation. The running pattern is most dependent on the joint angle motion in the sagittal plane, like hip flexion, compared to the joint angles moving in the horizontal or frontal plane. Furthermore, the sagittal plane is exposed to the highest range of motion, thereby being less exposed to noise and artifacts compared to the other planes. Thus it was chosen to use a sagittal camera view, to increase the possibility of accurate joint angle estimates in this plane.

5.2. *Light conditions*

In the movement laboratory, the light frequency of the ceiling lamps caused a frequent change in light condition in the video recordings. Thus, instead of using the ceiling lamps, each recording was made with an external spotlight placed approx. three meter from the subject with an angle of approx. 45^o to the sagittal plane of the runner, this placement was standardized through all trials to prevent any change in light condition and to minimize background variation in regard to the shadow of the runner.

5.3. *Parallax error*

The issue of parallax error is known to influence the estimation of object position [12]. Parallax error occurs when the measurement of an object's length is more or less than the true length because the viewing position is at an angle to the object of interest. Video-image systems often require calibration of field of view based on field measurements of certain geometric elements before the data collection process can be initiated. Most video-image systems provide internal algorithms for such a calibration [13]. However, when wanting to use simple smartphone recordings, the pose estimation may be influenced by parallax error. In the current study, the viewing angle between the camera and the subject was standardized as much as possible, without influencing the natural running pattern of the subject.

In the experimental setup, the camera was placed with a distance of 2,30 m to the treadmill at a height of 1,44 m in a line approx. perpendicular to the hip of the subject. Before reaching steady state, the runner was instructed through verbal feedback to run with his hip approx. perpendicular to the middle of the treadmill, which was marked by a line of tape on the treadmill side. Ideally this would ensure that the runner in each recording was positioned with the body center in the camera viewing center, thereby minimizing and standardizing the influence of parallax error on the pose landmark positions.

6. Pose Estimation joint angle comparison

The statistical parametric running model includes 104 different DOFs that can be constrained by Fourier coefficient input generated from MediaPipe Pose. However, several of these could not easily be processed from the single sagittal view. Furthermore, the running model has the advantage that it is capable of predicting the most plausible running pattern from a minimal trial and subject-specific input. When the model is determining many parameters based on a few, the accuracy of these few input parameters should be as high as possible, since it can be argued that an estimation based on inaccurate input would equally result in inaccurate output. Thus, subsequent to the final data collection, a comparison of joint angles generated through MediaPipe pose estimation and joint angles generated from marker based motion capture was made.

The joint angles investigated were selected based on two perspectives. Firstly, they should have a relevant influence on the running pattern, and secondly they should be equal to one of the 104 DOFs included in the running model, with a time series signal comparable to the corresponding DOF throughout the gait cycle. Thus, the elevation of the shoulder joints, and the flexion of the knee, hip and elbow joints were compared with the equivalent angles derived from marker based motion capture. The pose estimation angles were before the comparison transformed to 11 Fourier coefficients and subsequently transformed back to time series. The comparisons were based on Pearson's correlations and root mean square differences from the 51 included trials.

Table 1: The median (min - max) of the Pearson's correlation coefficients and root mean squared differences between the pose estimation joint angles and the marker-based reference.

Variable	Right knee	Left knee	Right hip	Left hip	Right shoulder	Left shoulder	Right elbow	Left elbow
r	0.970 (0.901 - 0.991)	0.979 (0.934 - 0.991)	0.982 (0.917 - 0.994)	0.972 (0.897 - 0.990)	0.452 (0.023 - 0.906)	0.384 (0.006 - 0.803)	0.347 (0.006 - 0.913)	0.600 (0.028 - 0.991)
RMSD	17.5 (9.9 - 22.9)	9.5 (6.2 - 14.1)	8.5 (3.8 - 17.3)	9.5 (2.7 - 15.0)	33.2 (22.9 - 59.5)	39.4 (20.8 - 64.8)	17.8 (11.9 - 38.1)	16.4 (8.1 - 31.1)

Based on the results shown in Table 1, the input parameters in the current study were selected. When wanting to predict running patterns, it may seem confusing that the plantar flexion was not investigated. Early in the study process, it was established from visual inspection of the landmark position and joint angle plots from the pose estimation that the ankle and toe landmark positioning was too inconsistent, to generate a valid plantar flexion parameter.

Looking at Pearson's correlations, it clearly shows that the hip and and knee joint angle estimation curve courses are closely comparable to the reference, while the shoulder and elbow joint angle estimates are inaccurate compared to the reference. Hence these four DOFs were excluded from the input parameters.

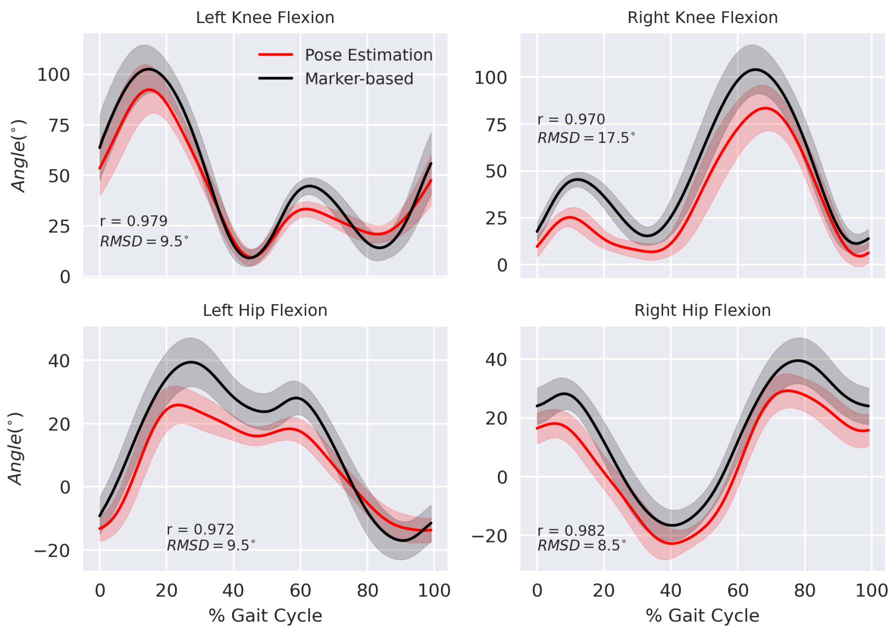

Figure 3: Left and right knee and hip flexion means and standard deviations from Pose Estimation and marker-based reference. The horizontal axis is normalized to 100% gait cycle, and the vertical axis is degrees.

The magnitude errors of the hip and knee do however differentiate from the marker-based reference. As shown in Fig 3, the Pose estimated joint angles tend to be below the reference. Thus, it was decided not to include a0 coefficients in Model-PE in the article. Furthermore, based on the variation shown in Fig 3, it was established that coefficient selection analysis was necessary, to ensure the right amount of constrained knee and hip angle coefficients in Model-PE. This was done to ensure that all model input was improving the running pattern estimation and not regressing it.

7. Segment length combination analysis

The total of 1224 predicted and constrained running parameters from the parametric running model were used to drive a simulation of a musculoskeletal model in the AnyBody Technology software similar to the one driven by the marker-based reference. A TrialData.any file containing information about the running speed, age, gender, anthropometry and 11 Fourier coefficients for all degrees of freedom were created for each trial. This file contains a kinematic Fourier driver for each degree of freedom, where a_0 and the five a and b coefficients are used to drive the model for one gait cycle. This enabled visual sanity checks of the estimated running pattern and anthropometrics through the model view function, see Fig 4. Through this visual inspection it was clear that constraining the pelvis width led to unrealistic model scaling, hence this parameter was not constrained.

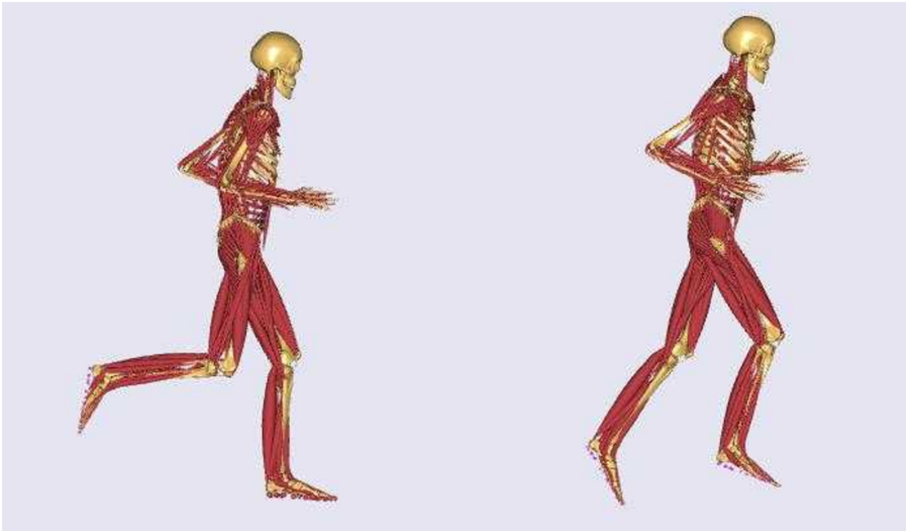

Figure 4: AnyBody Technology model view of two different running gait frames driven by parametric running model estimations.

When investigating the different segment combinations in the visual model representation, several different combinations could be constrained, without having notable influence on the representation. Including or excluding a specific segment length to the constrained model input has influence on the output estimation, where it may improve some parameter estimations while worsening others. For instance, trials showed that including lower and upper arm lengths to the constrained parameters improved the elbow flexion estimation, while worsening the plantar flexion estimation. Thus, a segment combination analysis was done, to ensure the overall best performing segment combination was being used in the study. Trials showed that the relative differences of the model performance when using different segment combinations were not notably affected by the number of coefficient pairs being constrained. Thus, all possible combinations of segment lengths were compared using one coefficient pair. Segments represented on both sides of the body were only constrained for the left side, because the parametric running model automatically assumes symmetry between the right and left body half. Hence, constraining the left upper arm to a specific length will equally constrain the right upper arm to the same length.

Table 2: The median (min - max) of the Pearson's correlation coefficients and root mean squared differences between the pose estimation driven vertical center of mass estimated by constraining different segment length combinations and the marker-based reference.

Variable	None	All	Legs and arms	Legs	Arms	Legs and feet	Legs and trunk height
r	0.961 (0.779 - 0.996)	0.850 (0.297 - 0.977)	0.957 (0.828 - 0.992)	0.969 (0.828 - 0.997)	0.946 (0.756 - 0.988)	0.969 (0.828 - 0.996)	0.880 (0.338 - 0.990)
RMSD	0.021 (0.008 - 0.049)	0.028 (0.012 - 0.043)	0.032 (0.005 - 0.061)	0.030 (0.005 - 0.055)	0.019 (0.007 - 0.058)	0.030 (0.005 - 0.055)	0.027 (0.011 - 0.044)

Unlike the majority of the different outputs from the parametric running model, vCoM is not restricted by a specific joint or segment location. Therefore, it became the decisive output parameter for the segment combination analysis. Table 2. shows a selection of the results from the segment combination analysis. The analysis is based on different combinations of constraining the segment length of the foot, shank, thigh, upper arm, lower arm, and trunk height. Furthermore, it was also investigated not constraining any of the segment lengths. The comparison showed highest vCoM median correlation when constraining the legs (shank and thigh lengths) or the leg and foot lengths, while it showed the lowest median RMSE when only constraining the arms (upper and lower arm lengths). However, constraining the leg lengths or leg and foot lengths resulted in a poor RMSD median value compared to other combinations. Oppositely, constraining the arms showed poor median correlation compared to other combinations. Whereas, without constraining any segment lengths, the model showed vCoM median correlation similar to constraining the legs and RMSD median similar to constraining the arms. Thus, it was chosen not to constrain any segment lengths.

This decision furthermore had the advantage that model-PE in a practical context became less complicated for use in field measurements.

8. The perspectives of TrackMan

This Study concludes that it is possible to generate results comparable to marker based motion capture, when utilizing a statistical parametric running model driven by joint angles from pose estimation. Our results show that this concept is capable of generating similar results to other studies investigating single or multi view gait analysis in running [14, 15, 16]. However, it is argued that the present pose estimation technology alone is not meeting the requirements of motion analysis in sports [17]. In the context of TrackMan using this concept to develop a parametric golf model, the results of the parametric running model driven by knee and hip kinematics should not vary significantly from the marker-based reference. Thus, we have investigated the performance of Model-PE further through a paired, one-dimensional, two-tailed Statistical parametric mapping (SPM) t-test with an alpha value of 5%, comparing it to the marker-based reference. The SPM as a statistical method, illustrates an intuitive understanding of where the significant differences occur over time of the gait cycle. It has the advantage that it analyzes time series data as continuous data, whereas traditional statistics often rely on discretized characteristics of the data [18].

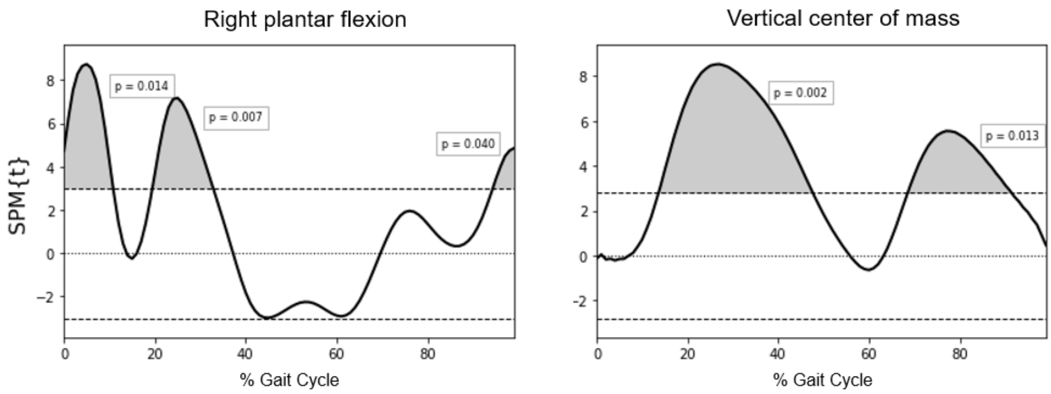

Figure 5: Result of SPM paired t-test for right plantar flexion (left) and vertical center of mass (right).

In Fig. 5 it is shown that the right plantar flexion angle and vCOM are significantly larger in Model-PE compared to the marker-based reference multiple times throughout the gait cycle. This is furthermore the scenario for all Model-PE output parameters. Differences were expected, since the validity of synthetically created movement kinematics will always be inferior to laboratory recorded data [1]. Thus, the estimations from Model-PE are subject to some degree of uncertainty, which is part of the tradeoff when simplifying the measurement approach. However, the results of Model-PE are generated based on simple 2D joint angle calculations for only the knee and hip joints derived from a simple smartphone recording. Hence, identifying other parameters essential for estimating the running pattern, could potentially be a way of further developing on the concept of the current study. Considering the simple method used in this study, one can expect a potential for developing a more complex or movement-specific joint angle calculation approach.

In the perspective of using pose estimation and a biomechanical model to estimate the movement of a golf swing, parameters that contain much information about the swing characteristics need to be identified, in order to obtain trial specificity of the estimations. Furthermore, to ensure realistic estimations when constraining these parameters, they need to be measured precisely. Therefore, the validity of a parametric golf model would rely on both an identification of important movement parameters and the potential of developing accurate pose estimation joint angle calculations of these parameters. In this context, the potential of developing a golf-specific pose estimation algorithm could be investigated, with the purpose of tracking specific important parameters precisely, since the parametric model would rely on minimal but valid input. Furthermore, when analyzing a golf swing, it must be considered that a camera view from the side might not be optimal, since a lot of the range of motion is better captured from the frontal plane. Another dimension for a statistical parametric golf model could be to use information from the Doppler radar system as model parameters. This could potentially improve the model estimation, since the estimation would not only be based on the swing kinematics but also the output of the swing.

References

[1] J. Rasmussen, M. E. Lund, R. P. Waagepetersen, Data-based parametric biomechanical models for cyclic motions, in: DHM 2020 - Proceedings of the 6th International Digital Human Modeling Symposium, Advances in Transdisciplinary Engineering, IOS Press, 2020, pp. 372–379. doi:10.3233/ATDE200045.

[2] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, L. Shao, Deep 3D human pose estimation: A review, Computer Vision and Image Understanding 210 (2021) 103225. doi:10.1016/j.cviu.2021.103225.
URL `https://linkinghub.elsevier.com/retrieve/pii/S1077314221000692`

[3] X. Chen, A. Pang, W. Yang, Y. Ma, L. Xu, J. Yu, Sportscap: Monocular 3d human motion capture and fine-grained understanding in challenging sports videos, International Journal of Computer Vision 129 (10) (2021) 2846–2864.

[4] L. Sigal, A. O. Balan, M. J. Black, Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, International journal of computer vision 87 (1) (2010) 4–27.

[5] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments, IEEE transactions on pattern analysis and machine intelligence 36 (7) (2013) 1325–1339.

[6] J. Hwang, S. Park, N. Kwak, Athlete pose estimation by a global-local network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 58–65.

[7] H. Qiu, C. Wang, J. Wang, N. Wang, W. Zeng, Cross View Fusion for 3D Human Pose Estimation, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), 2019, pp. 4341–4350. doi:10.1109/ICCV.2019.00444.
URL `https://ieeexplore.ieee.org/document/9008809/`

[8] Y. Desmarais, D. Mottet, P. Slangen, P. Montesinos, A review of 3D human pose estimation algorithms for markerless motion capture, Computer Vision and Image Understanding 212 (2021) 103275. doi:10.1016/j.cviu.2021.103275.
URL `https://linkinghub.elsevier.com/retrieve/pii/S1077314221001193`

[9] Google, MediaPipe Pose (2020).
URL `https://google.github.io/mediapipe/solutions/pose.html`

[10] S. D. Skejø, M. E. Lund, M. Stensvig, N. M. Kaae, J. Rasmussen, Running in circles: Describing running kinematics using Fourier series, Journal of Biomechanics 115 (2021) 110187. doi:10.1016/j.jbiomech.2020.110187.
URL `https://linkinghub.elsevier.com/retrieve/pii/S0021929020306114`

[11] D. Webb, W. A. Sparrow, Description of joint movements in human and non-human primate locomotion using Fourier analysis, Primates 48 (4) (2007) 277–292. doi:10.1007/s10329-007-0043-4.
URL `http://link.springer.com/10.1007/s10329-007-0043-4`

[12] J. Stephens, M. Bostjancic, T. Koskulitz, A Study on Parallax Error in Video Analysis, The Physics Teacher 57 (3) (2019) 193–195. doi:10.1119/1.5092485.
URL `http://aapt.scitation.org/doi/10.1119/1.5092485`

[13] Z. Z. Tian, M. D. Kyte, C. J. Messer, , Journal of Transportation Engineering 128 (3) (2002) 218–223. doi:10.1061/(ASCE)0733-947X(2002)128:3(218).
URL

[14] R. M. Kanko, E. K. Laende, E. M. Davis, W. S. Selbie, K. J. Deluzio, Concurrent assessment of gait kinematics using marker-based and markerless motion capture, Journal of Biomechanics 127 (2021) 110665. doi:10.1016/j.jbiomech.2021.110665.
URL `https://linkinghub.elsevier.com/retrieve/pii/S0021929021004346`

[15] J. Stenum, C. Rossi, R. T. Roemmich, Two-dimensional video-based analysis of human gait using pose estimation, PLOS Computational Biology 17 (4) (2021) e1008935. doi:10.1371/journal.pcbi.1008935.
URL `https://dx.plos.org/10.1371/journal.pcbi.1008935`

[16] D. Pagnon, M. Domalain, L. Reveret, Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness, Sensors 21 (19) (2021) 6530. doi:10.3390/s21196530.
URL `https://www.mdpi.com/1424-8220/21/19/6530`

[17] S. L. Colyer, M. Evans, D. P. Cosker, A. I. T. Salo, A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System, Sports Medicine - Open 4 (1) (2018) 24. doi:10.1186/s40798-018-0139-y.
URL `https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-018-0139-y`

[18] T. C. Pataky, Generalized n-dimensional biomechanical field analysis using statistical parametric mapping, Journal of biomechanics 43 (10) (2010) 1976–1982.