data

MDPI

Article

TRIPOD—A Treadmill Walking Dataset with IMU, Pressure-Distribution and Photoelectric Data for Gait Analysis

Justin Trautmann [1,*], Lin Zhou [1], Clemens Markus Brahms [2], Can Tunca [3], Cem Ersoy [3], Urs Granacher [2] and Bert Arnrich [1]

1 Digital Health-Connected Healthcare, Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany; lin.zhou@hpi.de (L.Z.); bert.arnrich@hpi.de (B.A.)
2 Division of Training and Movement Sciences, University of Potsdam, 14469 Potsdam, Germany; mbrahms@uni-potsdam.de (C.M.B.); urs.granacher@uni-potsdam.de (U.G.)
3 NETLAB, Department of Computer Engineering, Bogazici University, Istanbul 34342, Turkey; can.tunca@boun.edu.tr (C.T.); ersoy@boun.edu.tr (C.E.)
* Correspondence: justin.trautmann@student.hpi.de

Abstract: Inertial measurement units (IMUs) enable easy to operate and low-cost data recording for gait analysis. When combined with treadmill walking, a large number of steps can be collected in a controlled environment without the need of a dedicated gait analysis laboratory. In order to evaluate existing and novel IMU-based gait analysis algorithms for treadmill walking, a reference dataset that includes IMU data as well as reliable ground truth measurements for multiple participants and walking speeds is needed. This article provides a reference dataset consisting of 15 healthy young adults who walked on a treadmill at three different speeds. Data were acquired using seven IMUs placed on the lower body, two different reference systems (Zebris FDMT-HQ and OptoGait), and two RGB cameras. Additionally, in order to validate an existing IMU-based gait analysis algorithm using the dataset, an adaptable modular data analysis pipeline was built. Our results show agreement between the pressure-sensitive Zebris and the photoelectric OptoGait system (r = 0.99), demonstrating the quality of our reference data. As a use case, the performance of an algorithm originally designed for overground walking was tested on treadmill data using the data pipeline. The accuracy of stride length and stride time estimations was comparable to that reported in other studies with overground data, indicating that the algorithm is equally applicable to treadmill data. The Python source code of the data pipeline is publicly available, and the dataset will be provided by the authors upon request, enabling future evaluations of IMU gait analysis algorithms without the need of recording new data.

Keywords: inertial measurement unit; gait analysis algorithm; OptoGait; Zebris; data pipeline; public dataset

Citation: Trautmann, J.; Zhou, L.; Brahms, C.M.; Tunca, C.; Ersoy, C.; Granacher, U.; Arnrich, B. TRIPOD—A Treadmill Walking Dataset with IMU, Pressure-Distribution and Photoelectric Data for Gait Analysis. *Data* **2021**, *6*, 95. https://doi.org/10.3390/data6090095

Academic Editors: Aleksandr Ometov and Joaquín Torres Sospedra

Received: 4 May 2021
Accepted: 23 August 2021
Published: 26 August 2021

1. Introduction

Human motion tracking, and human gait analysis in particular, constitute well-established methods in sports biomechanics and rehabilitation. Athletes and coaches use gait analysis to increase or maintain performance (i.e., return to sport), whereas clinicians and patients use it to evaluate the severity of neurological or orthopedic conditions (e.g., Parkinson's disease) [1]. Gait analysis can involve overground as well as treadmill walking. Overground walking is more common in daily life. However, due to the constraints of gold-standard measurement systems, such as instrumented walkways or multi-camera setups, data collection is limited to a small number of consecutive steps. Furthermore, installation of safety measures for the participant or patient during overground walking can be cumbersome [2]. It is well known that the biomechanics of treadmill walking differ from that of overground walking [3]. Since treadmill walking is common in sports and rehabilitation, reliable gait analysis data from treadmill walking are needed. Treadmill walking offers the opportunity to record a large number of consecutive steps in a single

trial while maintaining controlled conditions of walking speed and inclination. Individuals are kept within a constrained observation volume, and weight-support safety mechanisms can be routinely installed.

Gait analysis tools often rely on camera setups with multiple cameras and reflective markers (e.g., Vicon® motion capture systems) or instrumented walkways (e.g., GAITRite® pressure-sensitive walkway) [4]. While these systems have been widely used for conventional gait analysis, they are expensive, bulky or hard to set up [5]. In contrast, inertial measurement units (IMUs) potentially provide the same level of accuracy for spatiotemporal gait parameters at a much lower cost and with less effort to install and operate it and therefore represent an ecologically valid alternative to conventional gait analysis systems.

IMUs usually consist of a triaxial accelerometer and a triaxial gyroscope in a small casing. They can be effortlessly attached to any body segment to conveniently and unobtrusively measure an individual's movements. They are commercially available in large varieties at reasonable prices [6]. Given these advantages, IMU-based motion analysis of the lower limbs and gait analysis has gained popularity over the past years [7]. Numerous approaches have been proposed to obtain gait parameters via 3D-trajectories of the foot and lower limbs by means of numerical integration combined with zero-velocity updates (ZUPT) applied during the stance phase [8–10]. In addition, several commercially available software tools, such as the GaitUp LAB (Gait Up, Lausanne, Switzerland) or RehaGait® (Hasomed GmbH, Magdeburg, Germany) [11], can provide gait parameters derived from IMU data but lack detailed algorithm disclosure. In the context of treadmill walking, previous studies have evaluated the validity of IMUs for gait analysis based on comparisons of features directly derived from angular velocity and linear acceleration [12–14], whereas methods that estimate 3D trajectories from IMU data recorded on a treadmill and use them to estimate gait parameters have received less attention.

Previous studies have demonstrated that a treadmill does not represent an ideal inertial frame of reference since energy is exchanged between the treadmill belt and the participant in various amounts, depending on the treadmill power, speed, gait phase, participants' body mass and other factors [15–17]. Therefore, trajectory estimation algorithms that have been developed for overground walking and implicitly rely on a perfect inertial frame of reference and zero velocity during the stance phase may potentially exhibit worse performance when used with IMU data that have been recorded on a treadmill. However, to the best of our knowledge, there exists no publicly available dataset and software framework that could be used to evaluate the performance of such algorithms with data acquired on a treadmill. As shown in Table 1, prior research has recorded IMU data on a treadmill but either lacks a reliable reference system [8,18], did not estimate the 3D foot trajectories [12,19] or did not make the data available for research [8,11].

In order to address these limitations, this article presents a dataset called TRIPOD—**Tr**eadmill, **I**MU, **P**edobarographic and Ph**o**toelectric **D**ataset—that is available upon request, consisting of treadmill walking data collected from 15 young healthy adults. More specifically, the dataset contains IMU data recorded from the lower body, as well as reference data from a pressure distribution measurement system and a photoelectric gait analysis system. As a use case, the performance of a foot trajectory estimation algorithm from the literature [9] was implemented and evaluated using a modular and extendable data processing pipeline. The Python code for loading the dataset and example implementations of the used algorithms are made available on GitHub (https://github.com/HPI-CH/TRIPOD). The participants' consent requires accessing the data only with legitimate scientific interest; therefore, the dataset is available upon request for scientific use via a contact form on Zenodo (https://doi.org/10.5281/zenodo.5070771).

Table 1. Related work.

Study	Study Cohort	IMU Location	Reference Data	3D Trajectories	Publicly Available
[12]	17 patients with Parkinson's disease	in out-sole	Vicon	No	Yes
[11]	22 healthy participants	lateral aspect	Zebris	unknown	No
[8]	11 healthy participants	instep	reflective markers and GoPro HD2	Yes	No
[19]	13 healthy participants	thigh and shank	instrumented treadmill and Vicon	No	Upon request
[18]	108 healthy participants	foot, shank, thigh and pelvis	no reference system	Yes	Yes
This study	15 healthy participants	instep, heel, shank and sacrum	OptoGait and Zebris	Yes	Upon request

2. Materials and Methods

2.1. Participants

The study cohort consisted of 15 young and healthy participants (8 males, 7 females). The sample characteristics are described in Table 2. The eligibility for this study was determined by using the Physical Activity Readiness Questionnaire (PAR-Q) [20], and the level of recent physical activity was assessed by the International Physical Activity Questionnaires (IPAQ). The demographic and anthropometric data is provided in the `MetaInfo.csv` file. Participants were asked to avoid strenuous exercise in the 24 h prior to data collection and to bring their own comfortable footwear and clothes. All participants provided written informed consent after the experimental procedures were explained. The study was approved by the ethics committee of the University of Potsdam (63/2020), and all experiments were conducted according to the latest revision of the declaration of Helsinki.

Table 2. Sample characteristics. (SD: standard deviation, PWS: preferred walking speed on the treadmill).

	Minimum	Mean ± SD	Maximum
Age (yr)	20	26.4 ± 3.7	34
Mass (kg)	53.5	69.7 ± 12.1	103
Height (cm)	157.5	176.2 ± 8.8	190
Leg length (cm)	76	86.2 ± 4.3	95
PWS (km/h)	3.1	3.9 ± 0.5	4.9

2.2. Recording Devices

The reference data were recorded with an OptoGait system (Microgate, Bolzano, Italy) and a FDM-THQ pressure distribution measurement system (zebris Medical GmbH, Isny, Germany), subsequently referred to as the "Zebris system", that were both integrated in a quasar® med treadmill (h/p/cosmos sports and medical GmbH, Nussdorf-Traunstein, Germany). The Zebris system consists of 10,240 closely aligned capacitive force sensors at a density of 1.4 sensors/cm^2, which are integrated in the treadmill right underneath the treadmill belt [21]. The OptoGait system consists of a transmitting and a receiving bar of infrared LEDs, which are integrated into the foot boards of the treadmill at a length of 125 cm. During data acquisition, the LED sensors measure contact times and positions with a resolution of 1.041 cm at up to 1000 Hz [22].

Both recording devices, as well as the treadmill itself, were connected to a computer running the respective recording and control software. Raw sensor data, as well as calculated gait parameters, were saved after each recording session. More specifically, the OptoGait software provides spatiotemporal gait parameters, such as stride length, stride time, swing time and stance time, automatically after data acquisition. The Zebris software is designed for pedobarographic analysis and only provides data for stance time. Other gait parameters are only available as the mean and standard deviation for a selected number of steps as part of the gait analysis report that can be produced by the Zebris software. This does not allow a comparison of individual strides. Therefore, in our study, other gait parameters were calculated subsequently from raw sensor data. Detailed descriptions of the Zebris data can be found in Appendix B. Nonetheless, the gait reports obtained from the Zebris system were generated and included in the dataset for potential future use cases.

IMU data were recorded using seven Physilog®5 IMUs (Gait Up, Lausanne, Switzerland). The IMU sensors are factory-calibrated, and the raw data quality has been confirmed in previous studies [6]. Therefore, no further custom calibrations were performed in this study. Data were stored locally on an internal SD card in the IMU device, transferred to a computer via USB and converted from the device's binary format using the Physilog® Research Toolkit. Recordings were started and stopped synchronously on all sensors by connecting the sensors via Bluetooth to the GaitUp smartphone app [23]. The sensors were attached with Velcro straps to the outer shank, attached to the shoelaces at an instep position and attached on the heel of the shoe on each leg. Additionally, one sensor was placed at the sacrum on the back of a hip belt. IMU positions are shown in Figure 1.

The Zebris system and the IMUs were configured to record at 128 Hz, and the OptoGait system was configured to record at 1000 Hz. The gyroscope range was 1000°/s, and the acceleration range was initially set to 8 g according to pilot recordings. However, the acceleration range had to be increased to 16 g after some participants exhibited unexpectedly large accelerations during the heel strike.

Additionally, two RGB cameras were used to record the lower body of the participants in frontal and sagittal planes in order to document the experiment.

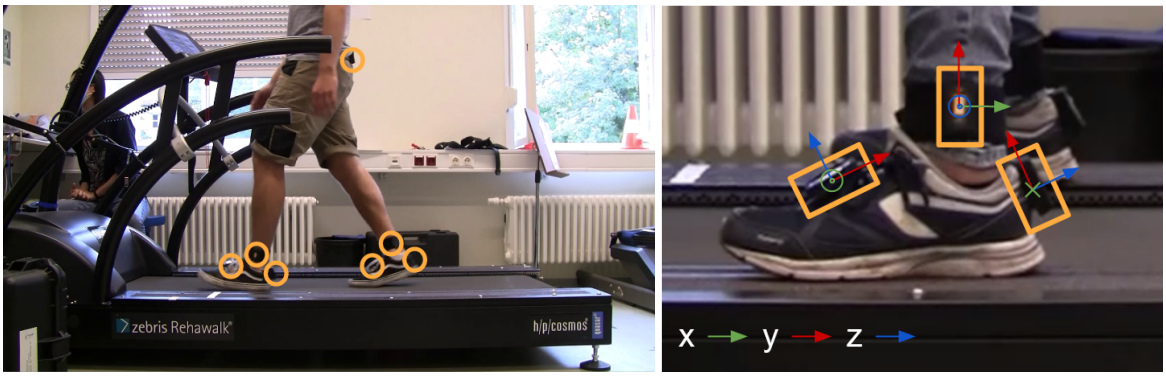

Figure 1. Experimental setup. IMU positions are highlighted in orange.

2.3. Experimental Protocol

Each participant performed three treadmill walking trials in a randomized sequence at their (i) preferred walking speed, (ii) 20% faster and (iii) 20% slower [24]. Each walking trial lasted for two minutes. The following protocol was implemented to prepare for data collection:

1. The participant received an explanation of the study from the instructor, signed the consent form and filled in the IPAQ and PAR-Q.
2. Body height, mass and leg length (between greater trochanter and lateral malleolus) were measured, and the IMU sensors were attached to the participant.
3. The participant was asked to walk for 6 min on the treadmill at a self-selected speed in order to get familiarized with treadmill walking [25].

4. The preferred walking speed (PWS) was determined by slowly increasing the treadmill speed until the participant reported that he/she was walking at their PWS. Then, the speed was substantially increased and lowered step by step until the participant reported PWS again. The final PWS was obtained by taking the average of the two reported PWS [24].
5. A 20% faster and 20% slower speed than PWS was calculated, and a random order of the three trials was determined.
6. All measurement systems were calibrated.

Subsequently, the following recording protocol was repeated three times for each participant:

1. The participant was asked to stand with legs apart on the foot boards of the treadmill.
2. The measurement systems were started, and the treadmill was set to the selected trial speed (PWS, PWS + 20%, or PWS − 20%).
3. The participant was asked to hold on the hand rails, step on the treadmill with the right foot first and release the handrails as soon as he/she felt comfortable. Stepping with the right foot is particularly important in order to obtain a well-identifiable signal for the initial contact from the IMUs and the reference systems that can be used to synchronize the different measuring systems.
4. The participants were asked to continue walking for two minutes from the moment of initial contact [26].
5. After two minutes, the treadmill was slowed down until it stopped, and the measurement systems were stopped.
6. If needed, participants were allowed to rest before reiterating the recording protocol for the remaining walking speed conditions.

3. Data

3.1. Preprocessing

The recorded data were exported from the respective recording tools and converted to manufacturer-independent open file formats. For the IMU and OptoGait data, the CSV format was chosen since the data can be well represented in a table. For the Zebris data, gzip compressed JSON files were chosen as the file format because of the nested structure of the data. The IMU data from the different sensors were synchronized using Physilog's built-in Bluetooth synchronization. All personal data, such as date and time of recording, were stripped from the files, and visible faces in the video recordings were blurred in order to preserve maximum anonymity for the participants and experimenters.

Since the data recording of the different measurement systems was started and stopped independently, all timestamps needed to be synchronized in order to match individual strides. The timestamp of initial contact with the treadmill belt can be directly obtained from the OptoGait and Zebris data since it is simply the first sample that is not null. However, determining the initial contact for the IMU data is more challenging because it does not correspond to the initial data sample but the first significant peak in the acceleration data from the right foot. When examining the raw data from different participants, it became apparent that a fixed threshold for the peak prominence cannot properly identify the initial contact for all participants. Therefore, a semi-automatic approach was implemented, in which the first significant peak in the acceleration data was identified automatically, and subsequently, the data plot was displayed to the user to allow manual correction. The resulting timestamps were written to the separate `SyncInfo.csv` file that is included in the dataset and can be used for synchronization.

3.2. Dataset

The dataset was organized according to participant, trial and measurement system, as shown in Figure 2. All data from one participant were stored in a folder named by a randomly generated, unique alias for the participant. Inside of the participant's folder,

the three trials were stored in three different folders containing sub-folders for each measurement system.

```
TRIPOD
|   MetaInfo.csv
|   SyncInfo.csv
|
+---Sub_AL
|   +---PWS
|   |   +---IMU
|   |   |         LF.csv
|   |   |         LH.csv
|   |   |         LS.csv
|   |   |         RF.csv
|   |   |         RH.csv
|   |   |         RS.csv
|   |   |         SA.csv
|   |   |
|   |   +---OptoGait
|   |   |         optogait.csv
|   |   |
|   |   \---Zebris
|   |       |     zebris_raw.json.gz
|   |       |     zebris_steps.json.gz
|   |       |
|   |       \---gait_report
|   |             parameters.csv
|   |              ...
|   |
|   +---PWS+20
|   |   + ...
|   |
|   \---PWS-20
|     + ...
|
+---Sub_BK
|   + ...
|
+ ...
```

Figure 2. Folder structure of the dataset.

Inside the `IMU` folder, there is one CSV file for each sensor (LF: left foot (instep), LS: left shank, LH: left heel, RF: right foot (instep), RS: right shank, RH: right heel, SA: sacrum). The `Zebris` folder contains two data files that correspond to the raw data export from the Zebris software and an aggregated data output that contains detailed information about each rollover cycle such as the time of initial contact and a maximum pressure distribution image. Additionally, the folder `gait_report` inside the `Zebris` folder contains the gait report as a collection of CSV files as exported from the Zebris software.

Aggregated gait parameters can be found in the `parameters.csv` file, and the documentation of the remaining files is available in the Zebris manual [21]. The OptoGait folder contains one CSV file that contains all the calculated gait parameters of the OptoGait software for each step.

The videos were organized in a separate folder tree of the same structure with one file for the side and one file for front view for each participant and trial, respectively, allowing the IMU and reference system dataset to remain small in size and independently usable from the videos. One participant was not included in the video dataset since permission for publication was denied.

Additionally, the root directory contains two CSV files with meta information such as participant characteristics, the order of trials and the questionnaire responses, as well as

one synchronization file that contains the timestamp of initial contacts in the IMU data for each participant and trial. The detailed file structure of JSON files containing the Zebris data is explained in Appendix B.

3.3. Data Processing Pipeline

A modular data processing pipeline was implemented in Python 3.7 to process the acquired data and evaluate the performance of existing trajectory estimation algorithms. The object-oriented design allows an easy exchange of the individual algorithm parts, making it convenient to combine trajectory estimation and gait event detection algorithms as well as different reference systems. The IMU data are first processed independently by a trajectory estimation algorithm and a gait event detection algorithm. Subsequently, the results are merged and matched stride by stride to the reference system's data. Finally, an evaluation part allows to generate correlation plots of the gait parameters from the IMU data and the reference system. Figure 3 shows an overview of the data processing pipeline.

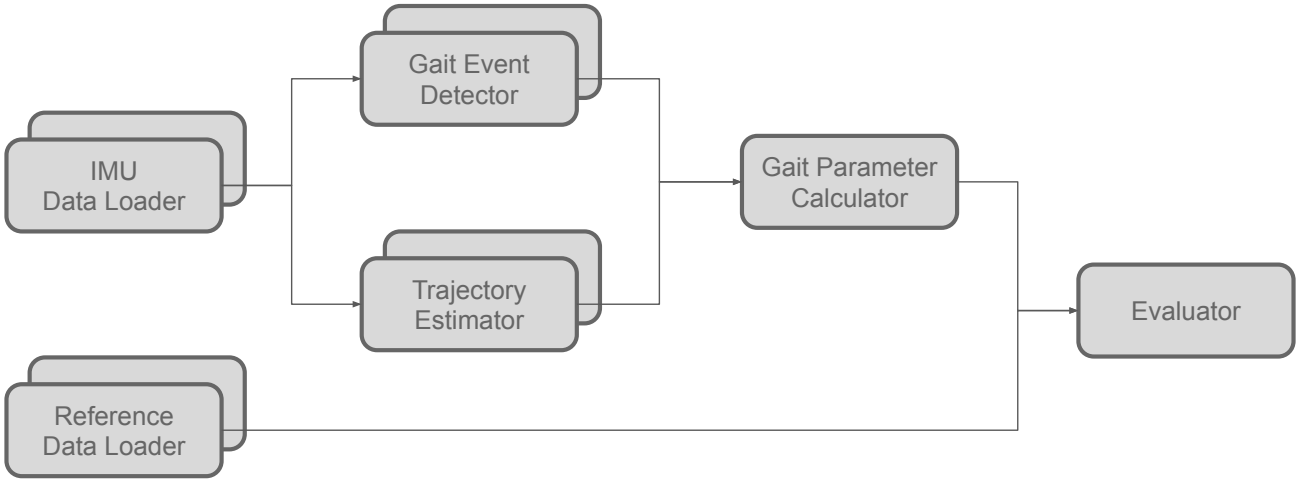

Figure 3. Data pipeline components and data flow. The data loader, reference loader, gait event detector and trajectory estimator were implemented as exchangeable classes, allowing the future implementation of multiple different algorithms and compatibility with different data formats.

As a use case, specific to the dataset, data loaders for Physilog IMU data and Zebris and OptoGait reference data were implemented. For the gait event detection and trajectory estimation, algorithms developed by Tunca et al. [9] were reimplemented. Tunca's gait event detection is based on an initial segmentation of the recording into strides based on minimal gyroscope energy, as presented by Skog et al. [27]. Within one stride, the foot's tilt angle reaches a minimum shortly after the foot leaves the ground and reaches a maximum shortly before it touches the ground. These two moments can subsequently be used as boundaries of search regions for the initial contact and foot off event. The foot off event is defined as the minimum of angular velocity of the foot prior to the the respective search region boundary, and the initial contact is defined as the first notable negative peak of angular velocity after the the respective search region boundary. The trajectory estimation algorithm from Tunca et al. is based on an error-state Kalman filter with zero-velocity updates during the stance phase and a following RTS smoother. During the swing phase, the position, velocity and orientation of the sensor is estimated using a dead reckoning approach, and during the stance phase, zero velocity of the foot is assumed, and the estimation is corrected, respectively, using a Kalman filter.

A Gait Parameter Calculator, as shown in Figure 3, is employed in order to calculate stride time, stride length, swing time and stance time from the estimated trajectory and the gait events. As shown in Figure 4, stride time, stance time and swing time can be directly inferred from the gait event detection in the top graph. In order to retrieve the

stride length, the timestamps of the initial contact events are used to identify the respective points of initial contact on the estimated trajectory in the bottom graph. Subsequently, the stride length is defined as the horizontal distance between the points corresponding to two subsequent initial contacts. Additional parameters, such as clearance, could not be validated due to vertical kinematic data not being present in the reference data.

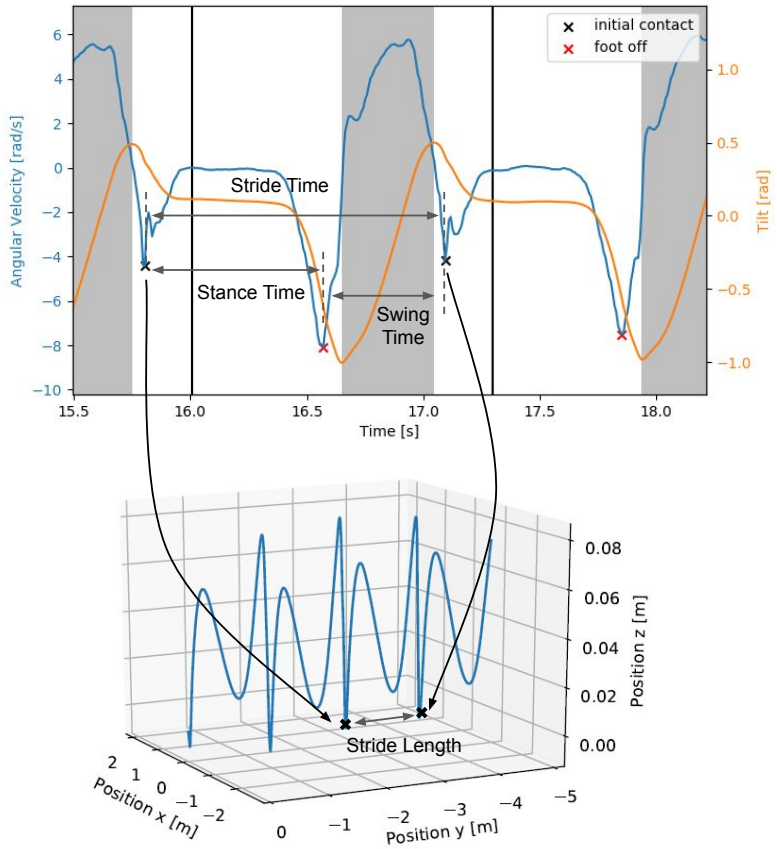

Figure 4. The estimation of gait parameters based on the combination of gait events and estimated 3D trajectory.

Spatial gait parameters were extracted from the Zebris raw data by displacing each pressure reading by the distance that results from the integration of the treadmill velocity up to the concerning sample and subsequently overlaying all of these pressure readings. The resulting pressure matrix represents the rollover imprints as if they were recorded on a stationary force plate and not on a treadmill.

Now, classical image processing techniques can be used to identify clusters. Figure 5 shows the tracking of foot prints over time. The actual imprint is color coded depending on the actual force on the belt. The displacement of each treadmill belt reading according to the integral of treadmill speed up to the respective point in time and the subsequent overlay of all samples leads to the data shown at the very right in Figure 5, leaving a trace of foot prints. Here, the lower edge of each cluster is detected, which is equivalent to detecting the very back end of the heel. The detected heel positions are attributed alternating to the right and left foot, illustrated by the green and blue lines.

In contrast to the Zebris system, OptoGait directly provides per stride gait parameters, requiring no raw data processing.

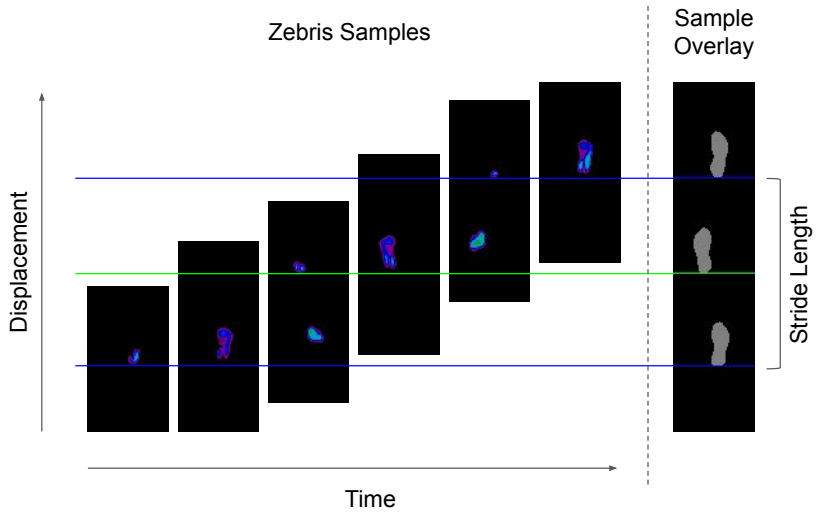

Figure 5. The evolution of Zebris readings over time and final heel detection. Green and blue lines: heel positions detected for the left and right foot, respectively.

4. Evaluation

Following the data acquisition, the performance of the reference systems as well as the trajectory estimation and gait detection algorithms was investigated.

In order to compare the results from the IMUs and the reference system for each individual stride, data from both sources need to be matched according to their timestamps. Even though all the systems are synchronized, the timestamp for each stride may be biased due to not perfect synchronization as well as subject to individual errors for each stride due to incorrect identification of the stride by either the reference system or the IMU algorithms. Thus, a nearest neighbor matching with a usual stride time as tolerance is employed. If strides cannot be matched with their counterpart in the other system, they are excluded from the analysis.

Subsequently, outliers were detected and excluded from further analysis. This is done by calculating a z-score (i.e., number of standard deviations above or below the mean value in a normal distribution) on the absolute difference between the corresponding values from the IMUs and the reference system and dropping all data records with a z-score greater than 3 (i.e., data that shows a greater difference between IMU and reference system than 99.7% of the data points). For individual test participants with very high standard deviation in the recorded data, filtering by z-score did not remove outliers sufficiently. Therefore, the data were filtered by a maximum difference between IMU and the reference system. Since the goal of the current analysis is to compare IMU data with reference data, it makes sense to filter outliers using the differences between the two data sources. In other use cases where only a single data source is available, alternative filtering techniques could be applied. For instance, in our IMU data, the main sources of error are merged or skipped strides, which result in extremely long (multiples of a normal stride) or short strides (close to zero), which could be filtered out using a z-score on the absolute values from the IMU system only.

The filtered data of all participants, trials and feet were merged and statistically analyzed. An ordinary least squares linear regression analysis was performed, yielding the parameters of the linear regression line, Pearson's r and the root-mean-square error. These parameters were subsequently used for the creation of correlation plots and Bland–Altman plots, as these plots allow assessing the inter-rater reliability between the IMU data and algorithms and the reference system.

4.1. Evaluation of the Reference Systems

As an initial data analysis step, the data from the OptoGait and Zebris system were evaluated against each other in the same way as IMU and reference data in order to validate their accuracy. By doing so, the results from previous studies could be replicated, indicating the correct functioning of both reference systems [28]. Figure 6 shows excellent agreement between the reference systems for stride length and stride time, indicating high-quality data for these gait parameters. However, Figure 7 shows significant bias between the two reference systems for stance and swing times. OptoGait measures notably longer stance phases and shorter swing phases compared to the Zebris system. This is a known issue inherent to the measurement systems that has been analyzed by Lee et al. [28]. The actual sensor placements of the OptoGait and the Zebris system are slightly above and below the treadmill belt, respectively, which causes the detected initial contact and foot off moments to deviate from the actual events and produce inaccurate stance and swing phases. Without a third accurate data source, the individual deviation of the Zebris and OptoGait system cannot be corrected. Therefore, in the following, only stride length and stride time are evaluated.

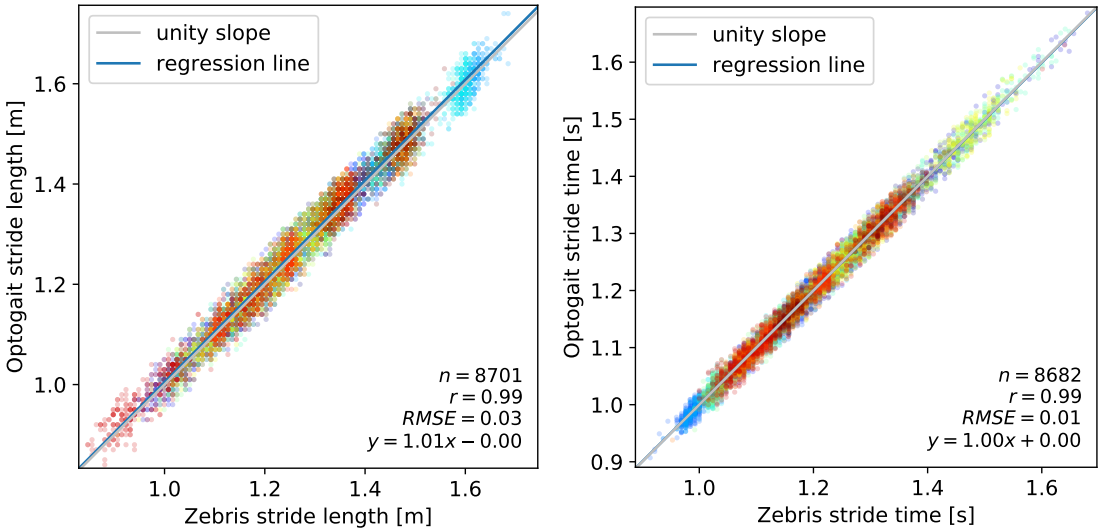

Figure 6. A comparison of stride length (**left**) and stride time (**right**) measured with the OptoGait and Zebris system. The different colors represent different participants.

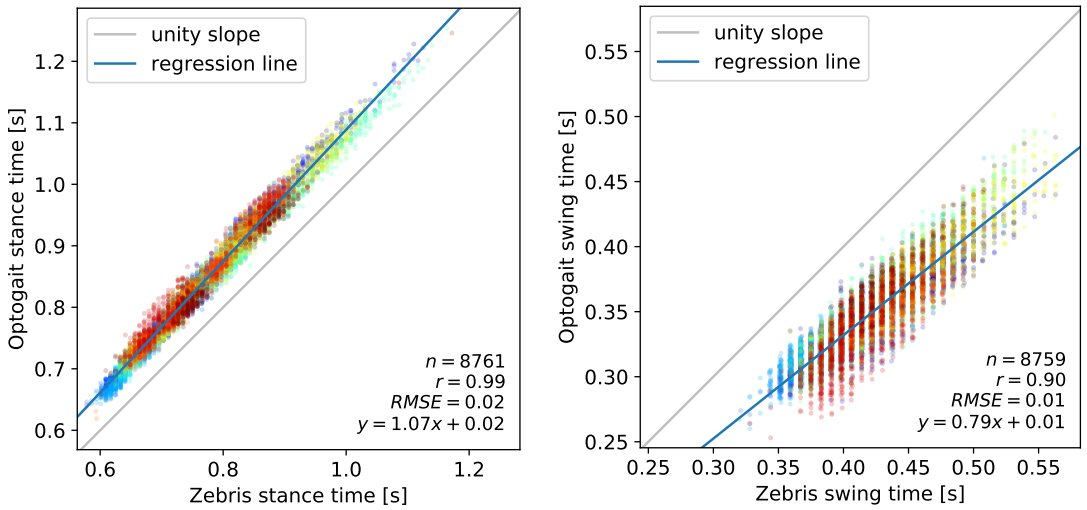

Figure 7. A comparison of stance time (**left**) and swing time (**right**) measured with the OptoGait and Zebris systems. The two measurement systems differ substantially due to different sensor placement—one above and one below the treadmill belt. The different colors represent different participants.

4.2. Evaluation of the IMU Gait Analysis Algorithms

The trajectory algorithm and gait event detection from Tunca et al. [9] were implemented and used to determine basic stride parameters from the IMU data of the right and left instep sensors. Since these algorithms were created for sensor data derived from instep sensors, only these were considered. However, data from different sensor placements are available in the dataset and could be used by other algorithms. The threshold that is required by the algorithms for determining stance phases from the gyroscope magnitude was selected by hand for each participant and trial. When comparing stride length and stride time estimates of the algorithm implemented in this study to the Zebris reference data, it is shown that the stride length estimation from the IMU data becomes less accurate as strides become large. In comparison, close agreement exists between the IMU gait event detection algorithm and the reference system, see Figure 8. This is probably due to the fact that larger steps are associated with larger absolute acceleration and gyroscope readings, which results in larger absolute errors in the raw data as well as in the processed data. As shown in Figure A1, the limits of agreement (LoA) between estimated and reference gait parameters are comparable to those reported for overground walking [9].

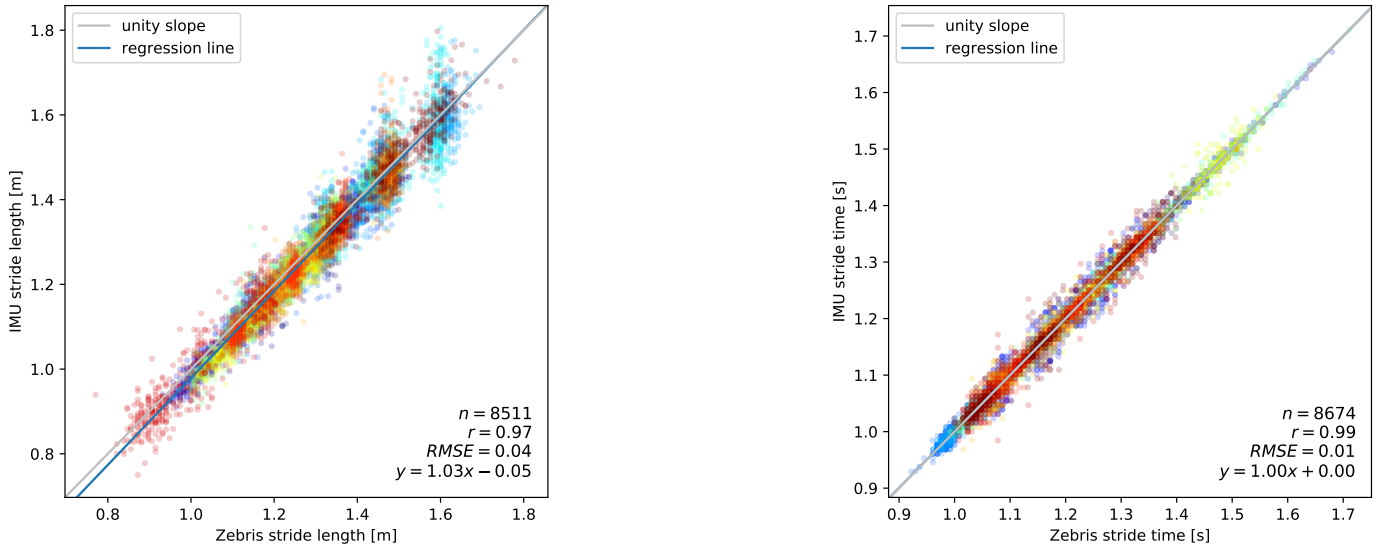

Figure 8. Comparison of the estimated and measured stride length (**left**) and stride time (**right**) from IMU data and the Zebris system. The different colors represent different trials.

When comparing stride length and stride time estimates to the measurements obtained by the OptoGait system, the aforementioned larger deviation for smaller strides cannot be observed (Figure 9). Instead, a slight bias over the range of stride lengths can be observed. Additionally, stride time estimation loses accuracy compared to the Zebris data. The 96% limit of agreement (LoA), as shown in Figure A2, is slightly lower for OptoGait than Zebris but still comparable to the data collected during overground walking. Tables 3 and 4 provide a direct comparison of the parameters shown in their respective plots.

Table 3. The quality of the estimated stride length. (RMSE: Root-mean-squared error, LoA: 96% Limit of Agreement).

	This Study		Tunca et al.
Reference System	**Zebris**	**OptoGait**	**Kinect v2**
Slope	1.03	1.00	0.97
Bias	−0.05	−0.02	0.02
RMSE (m)	0.04	0.05	0.05
LoA (m)	0.08	0.09	0.09

Table 4. The quality of the estimated stride time. (∗ swing time was evaluated in this study) (RMSE: Root-mean-squared error, LoA: 96% Limit of Agreement).

	This Study		Tunca et al. ∗
Reference System	**Zebris**	**OptoGait**	**Slow-Mo Camera**
Slope	1.00	0.99	1.00
Bias	0.00	0.02	0.00
RMSE (s)	0.01	0.02	0.02
LoA (s)	0.03	0.04	0.04

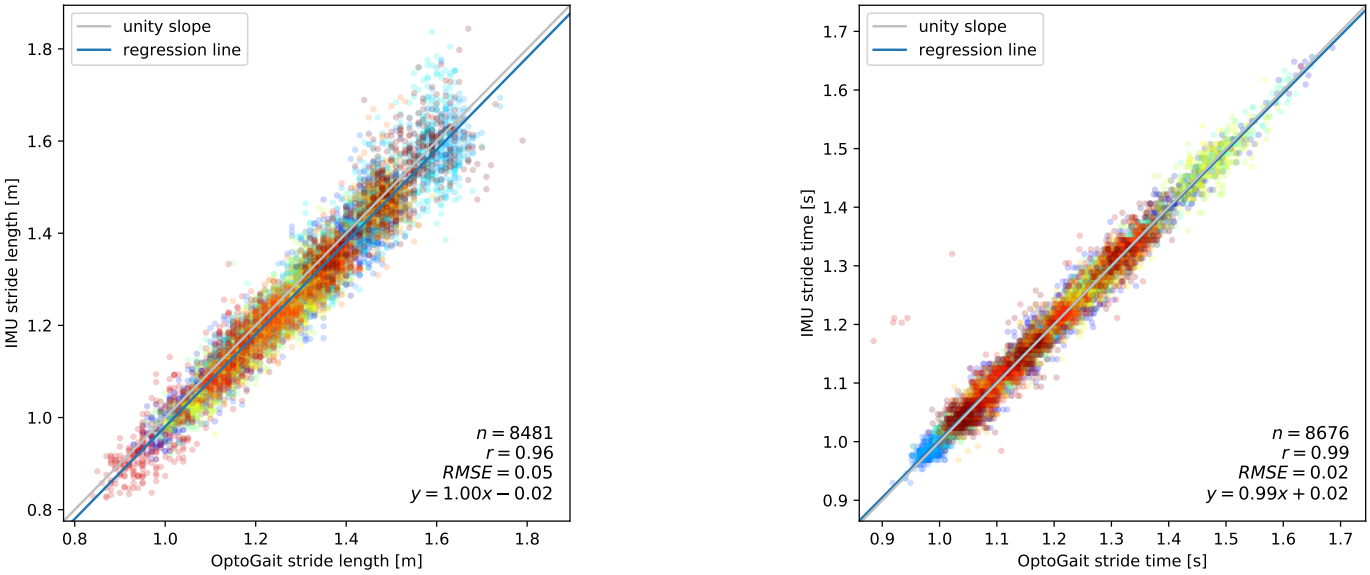

Figure 9. A comparison of the estimated and measured stride length (**left**) and stride time (**right**) from IMU data and the OptoGait system. The different colors represent different trials.

A comparative study with 22 participants by Tunca et al. [9] compared the performance of their respective algorithms to gait parameters obtained from a Microsoft Kinect v2 camera and a slow-motion camera for overground walking. The agreement between the reference system and the gait parameters estimates is comparable to the results of this study ($RMSE = 0.05$ m, $r^2 = 0.98$, $LoA = 0.09$ m for stride length, $RMSE = 0.02$ s, $r^2 = 0.95$, $LoA = 0.04$ s for swing time).

5. Discussion

This study presents a comprehensive dataset of IMU data, two reference systems, video data and meta data of 15 participants. We hypothesized that outcomes of the IMU gait analysis algorithm might suffer from the imperfect inertial frame on the treadmill. The results from this study were not in agreement with this hypothesis as the trajectory estimation algorithm and gait event detection algorithm used in this study produced similarly high-quality results for both treadmill and overground walking data (Tables 3 and 4). These results are also comparable to other state-of-the-art IMU gait analysis studies with respect to the LoA in stride length and time [10,29]. Consequently, the non-optimal inertial frame of a treadmill did not reduce the accuracy of the spatio-temporal gait parameter estimates for this particular algorithm. Nevertheless, other algorithms could potentially be more vulnerable when applied to treadmill IMU data.

Despite the high accuracy of stride time and stride length from the IMU system, we expected little agreement between the IMU system and either reference system for stance and swing time, since these two parameters differ significantly between the photoelectric OptoGait and pressure-sensitive Zebris system. When comparing these gait parameters

from the IMU system with both reference systems, as shown in Figures 10 and 11, it can be observed that the IMU system underestimates stance time and overestimates stride time. This was expected for the OptoGait system since it is installed above the treadmill belt and therefore detects contact with the belt slightly too early. The same trend was observed for the Zebris system because the sensors are installed underneath the treadmill belt and therefore detect contacts slightly too late. However, a slight delay in initial contact detection also exists for IMUs: given that the IMU analysis method needs a certain acceleration and change in angular velocity to detect a contact, this can only happen after the foot has touched the treadmill belt, and therefore, any detection of contact is also slightly delayed. The results show that this delay of the IMU system is larger than the delay of the Zebris system caused by the placement of the force sensors, explaining the underestimation of stance time by the IMU system. Consequently, deriving gait parameters from two different kind of gait events (e.g., swing and stance time from initial contact and foot off event) is not as reliable as when deriving gait parameters from gait events of the same kind (e.g., stride length and stride time from two consecutive initial contact events).

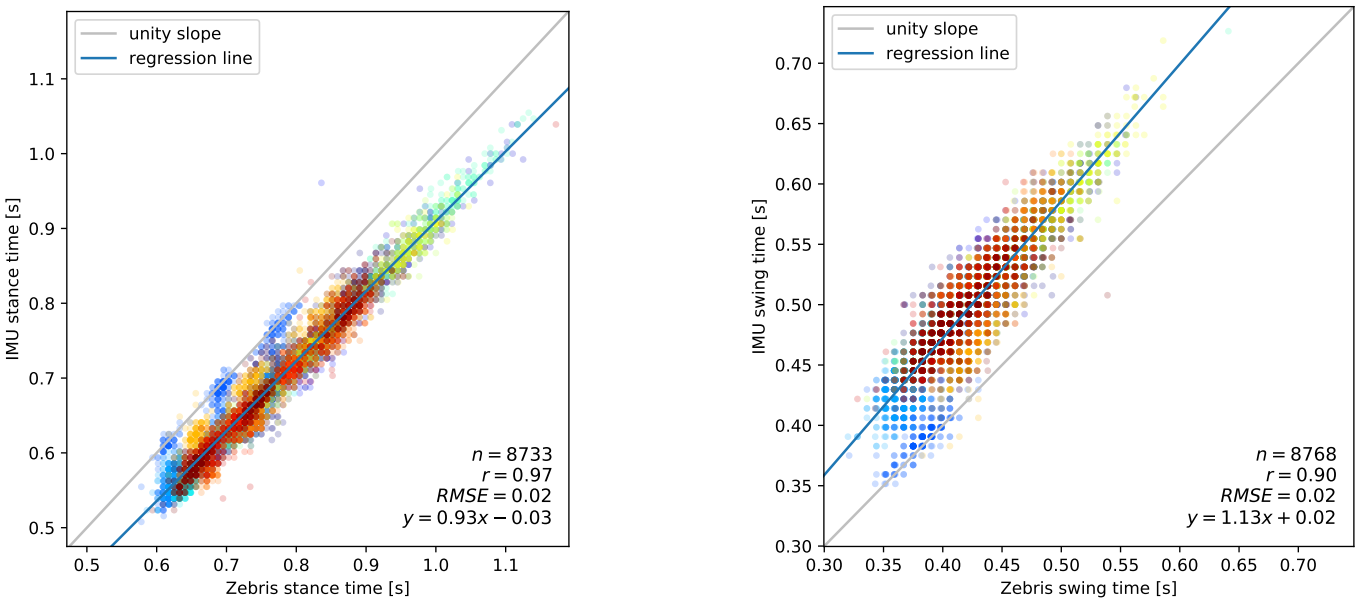

Figure 10. A comparison of the estimated and measured stance time (**left**) and swing time (**right**) from IMU data and the Zebris system. The different colors represent different trials.

The cohort of this study only included healthy individuals. Thus, it remains unclear if our methods produce similar results for individuals with pathological gait on the same treadmill setup. Nevertheless, Tunca et al. [9] showed that the algorithm performs well on data from overground pathological gait. If the conclusions can be transferred to treadmill walking, progress of rehabilitation or training could be tracked using IMU data and the study's data processing pipeline.

It is worth emphasizing that while the components of the data processing pipeline are easily exchangeable, the overall structure of the pipeline is fixed and limited to the present building blocks illustrated in Figure 3. Algorithms that rely on a different workflow (e.g., perform a combined estimation of trajectories and gait events or data-driven machine learning prediction of gait parameters) cannot be easily tested using this pipeline. However, the building blocks for data loading and evaluation could be potentially reused in a wide variety of pipeline design.

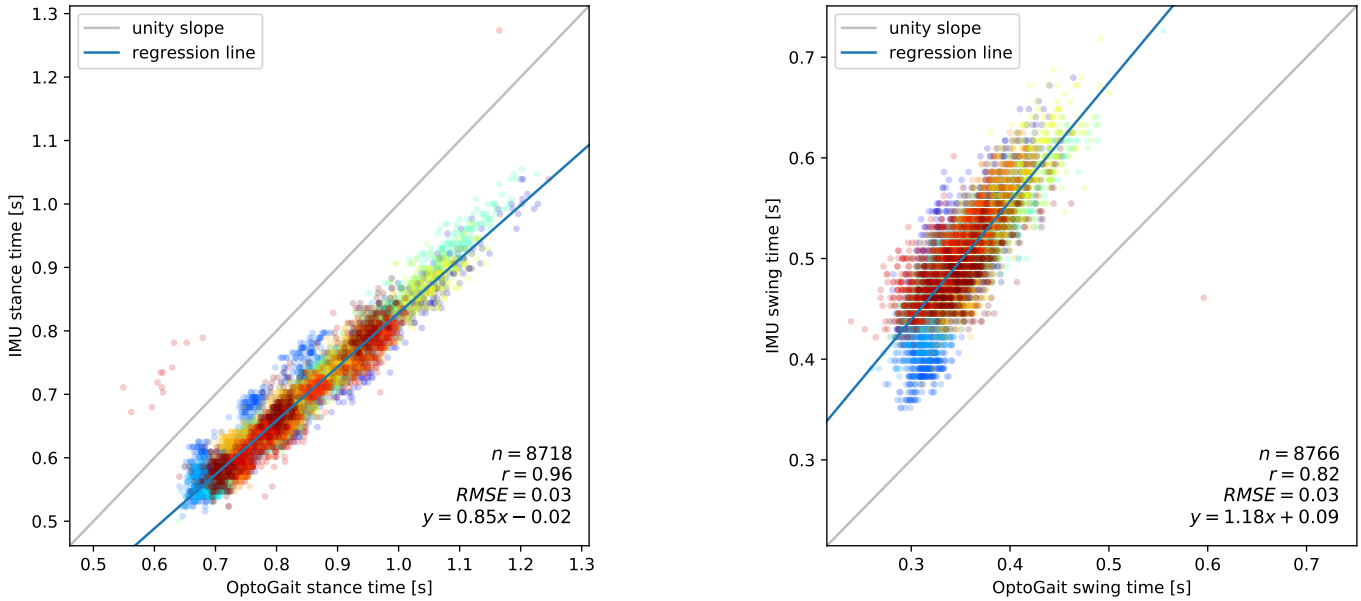

Figure 11. A comparison of the estimated and measured stance time (**left**) and swing time (**right**) from IMU data and the OptoGait system. The different colors represent different trials.

Furthermore, the presented dataset contains a multitude of data points that have not been analyzed in this study, such as data collected with IMUs placed on locations other than the instep, the raw pedobarographic data, various metadata, and videos. In this respect, our dataset presents a valuable resource for future research in the field of gait analysis.

Author Contributions: Conceptualization, J.T., L.Z. and B.A.; methodology, J.T., L.Z., C.M.B. and U.G.; software, J.T., C.T. and C.E.; validation, J.T. and L.Z.; formal analysis, J.T.; investigation, J.T. and L.Z.; resources, L.Z. and U.G.; data curation, J.T. and L.Z.; writing—original draft preparation, J.T.; writing—review and editing, L.Z., C.M.B., C.T., C.E., U.G. and B.A.; visualization, J.T.; supervision, L.Z., B.A. and U.G.; project administration, B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partly funded by the Federal Ministry of Education and Research of Germany in the framework of KI-LAB-ITSE (project number 01IS19066).

Institutional Review Board Statement: The study was approved by the ethics committee of the University of Potsdam (63/2020).

Informed Consent Statement: All participants provided written informed consent for this study.

Data Availability Statement: The dataset is available from the authors upon request for scientific purposes at https://doi.org/10.5281/zenodo.5070771. The source code used for analysis in this study can be found at https://github.com/HPI-CH/TRIPOD.

Acknowledgments: The authors would like to thank all participants who took part in this study, and Christiane Stielow for assisting with the data collection. The authors also acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
IMU	Inertial measurement unit
ZUPT	Zero-velocity update
PAR-Q	Physical Activity Readiness Questionnaire
IPAQ	International Physical Activity Questionnaires
PWS	Preferred walking speed
CSV	Comma separated values
JSON	JavaScript object notation
RMSE	Root-mean-square error
RTS	Rauch–Tung–Striebel
LoA	96% Limit of agreement

Appendix A. Additional Evaluation Plots

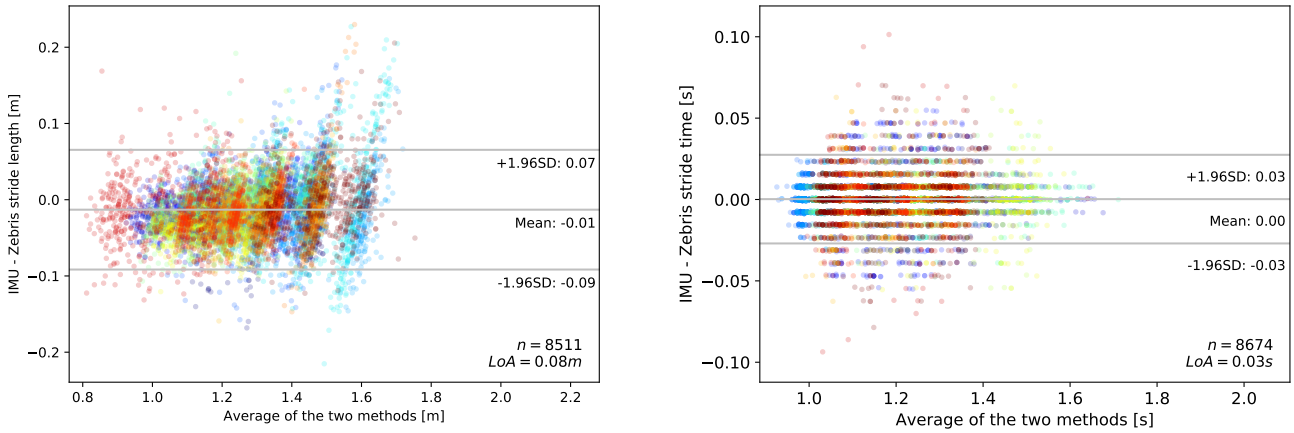

Figure A1. Bland–Altman plots of the estimated and measured stride length (**left**) and stride time (**right**) from IMU data and the Zebris system. The different colors represent different trials.

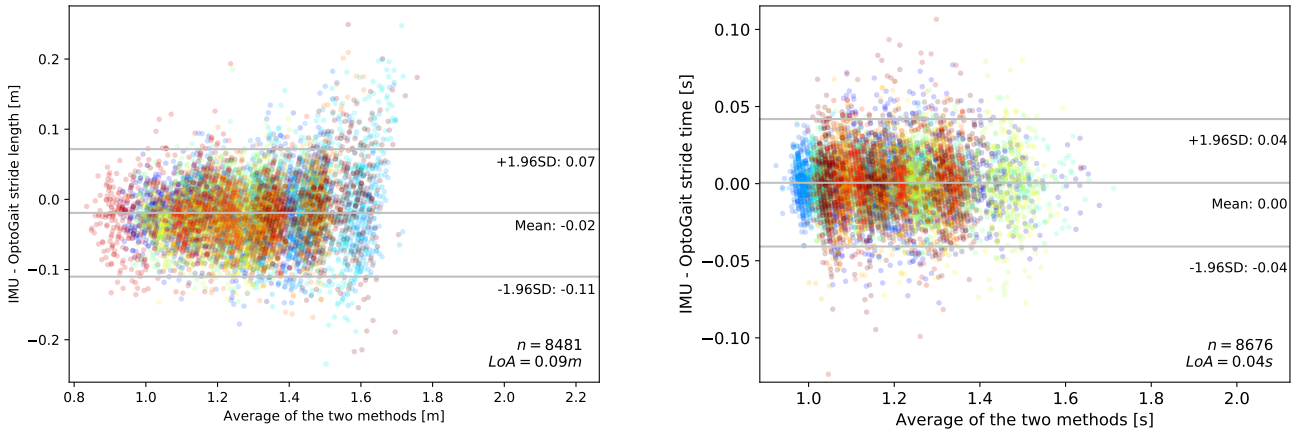

Figure A2. Bland–Altman plots of the estimated and measured stride length (**left**) and stride time (**right**) from IMU data and the OptoGait system. The different colors represent different trials.

Appendix B. JSON Data Format of Zebris Files

The content of the Zebris JSON file with raw data, as shown in Figure A3, contains the the following top-level keywords:

- `cell_count`: number of sensors along each axis of the device

- `cell_size`: size of each sensor in mm
- `frequency`: sampling frequency in Hz
- `pressure_unit`: literal of pressure unit
- `time_unit`: literal of time unit
- `velocity_unit`: literal of velocity unit
- `sample_count`: number of samples in the recording
- `begin`: timestamp of first sample (timestamps do not always start at zero)
- `samples`: list of all samples

The `cell_count` and `cell_size` members contain a `x` and `y` member, representing the respective coordinates.

The `samples` member consists of a list of objects. Each of these sample objects contains the following keywords:

- `origin`: origin of the data cell
- `size`: size of the data cell
- `pressure`: list of pressure values
- `velocity`: treadmill velocity

Since most of the time, the majority of pressure values are zero, the data are simply compressed by only providing data for the minimum rectangular area where values are not zero, as illustrated in Figure A4. These areas are defined by their origin and size. The `origin` and `size` members contain an `x` and `y` member that represents the respective coordinates. The list of pressure values is a flattened array in column-major (FORTRAN-style) order of the two-dimensional pressure data. In case all data are zero, the coordinates of the `origin` are `null`, the size is `0`, and the value of `pressure` is an empty array (`[]`).


```
{
      "cell_count": {
            "x": 64,
            "y": 160
      },
      "cell_size": {
            "x": 8.469,
            "y": 8.469
      },
      "frequency": 128,
      "pressure_unit": "N/cm\u00b2",
      "time_unit": "s",
      "velocity_unit": "m/s",
      "sample_count": 20947,
      "begin": 8.469,
      "samples": [
      {
            "origin": {
                  "x": 37,
                  "y": 93
            },
            "size": {
                  "x": 4,
                  "y": 5
            },
            "pressure": [
                  0.0, 2.2, 2.2, 0.0,
                  0.0, 3.8, 4.3, 1.6,
                  1.6, 6.5, 5.9, 3.2,
                  2.2, 3.8, 2.2, 0.0,
                  0.0, 1.1, 0.0, 0.0
            ],
            "velocity": 1.29
      },
      ...
      ]
}
```


Figure A3. An example file structure of a .json.gz files with Zebris data.

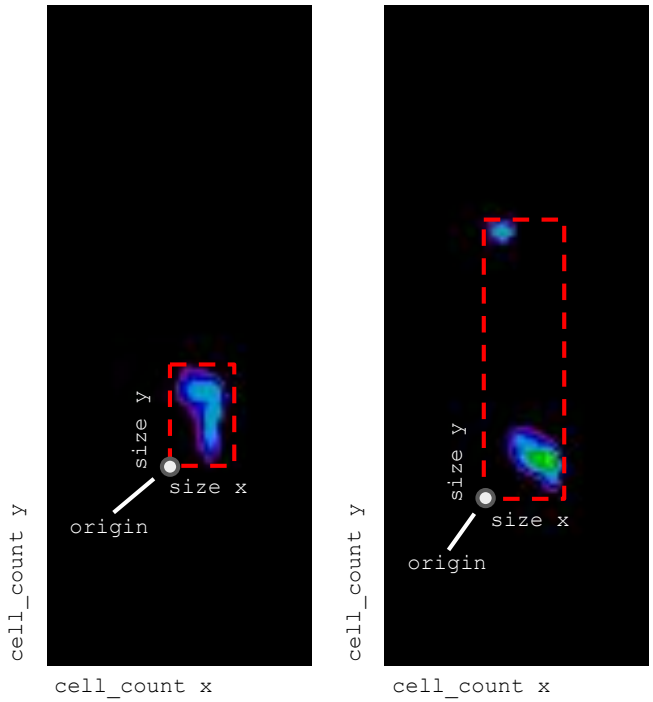

Figure A4. Illustration of pedobarographic data format for two particular data samples.

The content of the Zebris JSON file with aggregated data contains the maximum pressure distribution and rollover samples for each rollover cycle. This is referred to as `event`. The Zebris software calculates the rollover samples presumably by shifting and overlaying raw data samples according to the treadmill speed. The keywords `cell_count`, `cell_size` and `frequency` are equivalent to their counterpart in the raw data file. The remaining keywords are:

- `unit`: literal of pressure unit
- `event_count`: number of rollover cycles
- `events`: list of event objects

Each event object represents one rollover. It contains the following keys:

- `begin`: timestamp of first contact
- `end`: timestamp of last contact
- `side`: "left" or "right" indicating the considered foot
- `heel`: heel position with `x` and `y` member
- `toe`: toe position with `x` and `y` member
- `maximum`: maximum pressure distribution, analogue to one sample in the raw data file
- `rollover`: rollover object

A rollover object contains a series of samples that are analogous to the samples in the raw data file. However, since the frame for a rollover is not the sensor array but a dynamic window that encloses the footprint, the frame size is defined via the keyword `size`, and the number of samples during rollover is provided by the keyword `sample_count`.

References

1. Winter, D.A. *Biomechanics and Motor Control of Human Movement*; John Wiley & Sons: Hoboken, NJ, USA, 2009. [CrossRef]
2. Parijat, P.; Lockhart, T.E. Effects of moveable platform training in preventing slip-induced falls in older adults. *Ann. Biomed. Eng.* **2012**, *40*, 1111–1121. [CrossRef]
3. Riley, P.O.; Paolini, G.; Della Croce, U.; Paylo, K.W.; Kerrigan, D.C. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. *Gait Posture* **2007**, *26*, 17–24. [CrossRef] [PubMed]
4. Webster, K.; Wittwer, J.; Feller, J. Validity of the GAITRite walkway system for the measurement of averaged and individual step parameters of gait. *Gait Posture* **2006**, *22*, 317–21. [CrossRef]
5. Vicon Motion Systems Ltd. Vicon Vantage Quick Start Guide. 2020. Available online: https://docs.vicon.com/download/attachments/107483985/ViconVantageQuickStart.pdf (accessed on 19 January 2020).
6. Zhou, L.; Fischer, E.; Tunca, C.; Brahms, C.M.; Ersoy, C.; Granacher, U.; Arnrich, B. How We Found Our IMU: Guidelines to IMU Selection and a Comparison of Seven IMUs for Pervasive Healthcare Applications. *Sensors* **2020**, *20*, 4090. [CrossRef] [PubMed]
7. Weygers, I.; Kok, M.; Konings, M.; Hallez, H.; De Vroey, H.; Claeys, K. Inertial Sensor-Based Lower Limb Joint Kinematics: A Methodological Systematic Review. *Sensors* **2020**, *20*, 673. [CrossRef] [PubMed]
8. Ferrari, A.; Ginis, P.; Hardegger, M.; Casamassima, F.; Rocchi, L.; Chiari, L. A Mobile Kalman-Filter Based Solution for the Real-Time Estimation of Spatio-Temporal Gait Parameters. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 764–773. [CrossRef]
9. Tunca, C.; Pehlivan, N.; Ak, N.; Arnrich, B.; Salur, G.; Ersoy, C. Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders. *Sensors* **2017**, *17*, 825. [CrossRef]
10. Hori, K.; Mao, Y.; Ono, Y.; Ora, H.; Hirobe, Y.; Sawada, H.; Inaba, A.; Orimo, S.; Miyake, Y. Inertial Measurement Unit-Based Estimation of Foot Trajectory for Clinical Gait Analysis. *Front. Physiol.* **2019**, *10*, 1530. [CrossRef]
11. Donath, L.; Faude, O.; Lichtenstein, E.; Nüesch, C.; Mündermann, A. Validity and reliability of a portable gait analysis system for measuring spatiotemporal gait characteristics: Comparison to an instrumented treadmill. *J. Neuroeng. Rehabil.* **2016**, *13*, 6. [CrossRef]
12. Lee, M.; Youm, C.; Jeon, J.; Cheon, S.M.; Park, H. Validity of shoe-type inertial measurement units for Parkinson's disease patients during treadmill walking. *J. Neuroeng. Rehabil.* **2018**, *15*, 38. [CrossRef]
13. Yang, S.; Li, Q. IMU-based ambulatory walking speed estimation in constrained treadmill and overground walking. *Comput. Methods Biomech. Biomed. Eng.* **2012**, *15*, 313–322. [CrossRef]
14. Alam, M.N.; Khan Munia, T.T.; Fazel-Rezai, R. Gait speed estimation using Kalman Filtering on inertial measurement unit data. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2017**, *2017*, 2406–2409. [CrossRef] [PubMed]
15. Savelberg, H.H.C.M.; Vorstenbosch, M.A.T.M.; Kamman, E.H.; Weijer, J.G.W.v.d.; Schambardt, H.C. Intra-stride belt-speed variation affects treadmill locomotion. *Gait Posture* **1998**, *7*, 26–34. [CrossRef]
16. Sloot, L.; van der Krogt, M.; Harlaar, J. Energy exchange between subject and belt during treadmill walking. *J. Biomech.* **2014**, *47*, 1510–1513. [CrossRef] [PubMed]
17. Tielke, A.; Ahn, J.; Lee, H. Non-ideal behavior of a treadmill depends on gait phase, speed, and weight. *Sci. Rep.* **2019**, *9*, 12755. [CrossRef] [PubMed]
18. Loose, H.; Lindström Bolmgren, J. GaitAnalysisDataBase—Short Overview. 2019. Available online: http://gaitanalysis.th-brandenburg.de/static/files/GaitAnalysisDataBaseShortOverview.pdf (accessed on 23 April 2020).
19. Grimmer, M.; Schmidt, K.; Duarte, J.E.; Neuner, L.; Koginov, G.; Riener, R. Stance and Swing Detection Based on the Angular Velocity of Lower Limb Segments During Walking. *Front. Neurorobot.* **2019**, *13*, 57. [CrossRef] [PubMed]
20. Warburton, D.E.; Jamnik, V.K.; Bredin, S.S.; McKenzie, D.C.; Stone, J.; Shephard, R.J.; Gledhill, N. Evidence-based risk assessment and recommendations for physical activity clearance: An introduction. *Appl. Physiol. Nutr. Metab.* **2011**, *36*, S1–S2. [CrossRef] [PubMed]
21. zebris Medical GmbH. FDM-T Technical Specifications and User Manual. 2019. Available online: https://www.zebris.de/fileadmin/Editoren/zebris-PDF-Manuals/Medizin/Hardware/Aktuelle_Version/FDM-T_Hardware-Manual_Med_191204_en.pdf (accessed on 22 January 2020).
22. Microgate srl. OptoGait User Manual. Available online: http://www.optogait.com/optogaitportal/media/manuals/manual-en.pdf (accessed on 22 January 2020).
23. Gait Up SA. Physilog5 User Manual. 2018. Available online: https://research.gaitup.com/wp-content/uploads/2020/09/P5_Instructions-for-Use_V1.2.6.pdf (accessed on 22 January 2020).
24. Jordan, K.; Challis, J.H.; Newell, K.M. Walking speed influences on gait cycle variability. *Gait Posture* **2007**, *26*, 128–134. [CrossRef] [PubMed]
25. Meyer, C.; Killeen, T.; Easthope, C.S.; Curt, A.; Bolliger, M.; Linnebank, M.; Zörner, B.; Filli, L. Familiarization with treadmill walking: How much is enough? *Sci. Rep.* **2019**, *9*, 5232. [CrossRef] [PubMed]
26. Kribus-Shmiel, L.; Zeilig, G.; Sokolovski, B.; Plotnik, M. How many strides are required for a reliable estimation of temporal gait parameters? Implementation of a new algorithm on the phase coordination index. *PLoS ONE* **2018**, *13*, e0192049. [CrossRef]
27. Skog, I.; Handel, P.; Nilsson, J.; Rantakokko, J. Zero-Velocity Detection—An Algorithm Evaluation. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2657–2666. [CrossRef] [PubMed]

28. Lee, M.M.; Song, C.; Lee, K.; Shin, D.; Shin, S. Agreement between the spatio-temporal gait parameters from treadmill-based photoelectric cell and the instrumented treadmill system in healthy young adults and stroke patients. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2014**, *20*, 1210–1219. [CrossRef]
29. Yeo, S.S.; Park, G.Y. Accuracy Verification of Spatio-Temporal and Kinematic Parameters for Gait Using Inertial Measurement Unit System. *Sensors* **2020**, *20*, 1343. [CrossRef] [PubMed]