Heliyon 10 (2024) e26227

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Supervised learning for improving the accuracy of robot-mounted 3D camera applied to human gait analysis

Diego Guffanti [a,b], Alberto Brunete [c,*], Miguel Hernando [c], David Álvarez [d], Javier Rueda [e], Enrique Navarro [e]

[a] Centro de Investigación en Mecatrónica y Sistemas Interactivos - MIST, Universidad Indoamérica, Av. Machala y Sabanilla, 170103, Quito, Ecuador

[b] Universidad UTE, Av. Mariscal Sucre, Quito, 170129, Ecuador

[c] Centre for Automation and Robotics (CAR UPM-CSIC), Universidad Politécnica de Madrid, 28012 Madrid, Spain

[d] Department of Electrical, Electronic and Automation Engineering and Applied Physics, ETSIDI, Universidad Politécnica de Madrid, 28012 Madrid, Spain

[e] Department of Human Health and Performance, Faculty of Sports Sciences, Universidad Politécnica de Madrid, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
3D camera
Mobile robot
Gait analysis
Machine learning

ABSTRACT

Background and Objective: the use of 3D cameras for gait analysis has been highly questioned due to the low accuracy they have demonstrated in the past. The objective of the study presented in this paper is to improve the accuracy of the estimations made by robot-mounted 3D cameras in human gait analysis by applying a supervised learning stage. Methods: the 3D camera was mounted in a mobile robot to obtain a longer walking distance. This study shows an improvement in detection of kinematic gait signals and gait descriptors by post-processing the raw estimations of the camera using artificial neural networks trained with the data obtained from a certified Vicon system. To achieve this, 37 healthy participants were recruited and data of 207 gait sequences were collected using an Orbbec Astra 3D camera. There are two basic possible approaches for training and both have been studied in order to see which one achieves a better result. The artificial neural network can be trained either to obtain more accurate kinematic gait signals or to improve the gait descriptors obtained after initial processing. The former seeks to improve the waveforms of kinematic gait signals by reducing the error and increasing the correlation with respect to the Vicon system. The second is a more direct approach, focusing on training the artificial neural networks using gait descriptors directly. Results: the accuracy of the 3D camera to objectify human gait was measured before and after training. In both training approaches, a considerable improvement was observed. Kinematic gait signals showed lower errors and higher correlations with respect to the ground truth. The accuracy of the system to detect gait descriptors also showed a substantial improvement, mostly for kinematic descriptors rather than spatio-temporal. When comparing both training approaches, it was not possible to define which was the absolute best. Conclusions: supervised learning improves the accuracy of 3D cameras but the selection of the training approach will depend on the purpose of the study to be conducted. This study reveals the great potential of 3D cameras and encourages the research community to continue exploring their use in gait analysis.

* Corresponding author.
E-mail addresses: diego.guffanti@ute.edu.ec (D. Guffanti), alberto.brunete@upm.es (A. Brunete).

https://doi.org/10.1016/j.heliyon.2024.e26227
Received 18 January 2023; Received in revised form 8 September 2023; Accepted 8 February 2024

Available online 13 February 2024

1. Introduction

3D cameras have been widely used for gait analysis applications. Most studies claim that these sensors are not accurate enough to be used for gait analysis [1,2]. The low accuracy is more noticeable when analyzing gait kinematics than spatio-temporal parameters [3]. There are several factors that can influence the accuracy of these sensors: light conditions, sensor position, occlusion of different parts of the body, among others [4]. However, the main advantage of these sensors is undoubtedly their low cost and mode of operation. As they are markerless systems, they greatly facilitate the work of clinicians when analyzing gait and avoid disturbing patients when placing markers.

Machine Learning (ML) is an alternative that has opened the gap to improve the accuracy of 3D cameras. However, current approaches have only applied ML together with 3D cameras for the classification of normal and pathological gait [5], gait recognition [6], or for the classification of different pathologies [7]. On the contrary, they have not pursued an enhancement of the accuracy of these sensors. This is highly relevant for clinicians dealing with gait pathologies. For this reason, as a contribution to current approaches, this study aims to improve the accuracy of the estimated kinematic gait signals (joint angles) and gait descriptors when using 3D cameras. For this purpose, artificial neural networks (ANNs) are used. ANNs allow the system to learn from a gold standard system through supervised learning. It should be emphasized that in this study, ANNs do not learn a specific gait pattern. On the contrary, this study attempts to correct the measurement errors of the 3D camera based on the examples provided by the gold standard system. To achieve this, a windowing process is applied to the data and each window represents a training sample. This means that the ANN never learns the gait pattern as such. Consequently, it is possible to apply the training result to the analysis of new gait patterns obtaining better estimations.

In this study, neural networks were trained based on two criteria: training based on kinematic gait signals, and training based on gait descriptors. The former seeks to improve the waveforms of kinematic signals in sagittal, frontal and transverse planes, by reducing the error and increasing the correlation with respect to a Vicon system. The second is a more direct approach, focusing on training the networks using gait descriptors (like step width, stride length, cadence, max. flexions, max. extensions and rotations at the joints, among other descriptors) directly. In this second approach, 23 most important descriptors for distinguishing pathological from normal gait were selected for the analysis. This set of descriptors was obtained according to the recommendations of clinicians Molina and Carratalá [8].

The outline of this paper is organized as follows. Section 2 presents a state-of-the-art review of current methods applied to improve the accuracy of 3D cameras. Section 3 is dedicated to the methodology. This section explains the data collection protocol, as well as the process followed for the definition of joint kinematics. Section 4 addresses the two training approaches and the system accuracy achieved in each. Section 5 is focused on the comparison of both training approaches. Section 6 discusses the results of this study by comparing them with the current approaches. Finally, Section 7 concludes this study.

2. State of the art

The accuracy of 3D cameras in gait analysis has been a highly discussed topic. For example, Springer and Seligmann [3] presented a review of 12 studies that assessed gait analysis with a Microsoft Kinect sensor and a gold standard system. Results of this review indicated good validity for only some spatio-temporal parameters, and not enough validity for gait kinematics variables. For this reason, different approaches have pursued to improve the accuracy of 3D cameras in gait analysis. These approaches have covered, among others, the fusion of 3D cameras with inertial sensors [9], fusion of multiple 3D cameras [10–12], or the use of regression equations [13] to increase accuracy in the detection of foot-off and foot-contact events. In [12], Müller et al. applied a multi-camera configuration that was intended to improve the estimation of spatio-temporal parameters. In the study developed by Matthew et al. [14], the low accuracy of 3D cameras is attributed to the lack of a gait model for the data retrieved by the sensor, therefore, the authors propose the improvement of the estimations applying a new model based on rigid bodies. Nichols et al. [15] presented another approach based on retro-reflective markers. The researchers found that retro-reflective markers could be used to enhance the motion capture process with 3D cameras.

Another approach has been the fusion with inertial sensors. In the study presented by Alizadegan and Behzadipour [16], the authors proposed a new method to improve accuracy and real-time performance of inertial joint angle estimation for upper limb rehabilitation applications. The position measurements retrieved from a Kinect sensor were used to correct for the sensor-to-segment misalignment of the inertial sensors. Also in the study presented by Destelle et al. [17] the authors built an improved skeleton using information fused from a Kinect sensor and nine inertial sensors fixed to the subject's forearms, arms, thighs, shanks and chest. This method proved to obtain more accurate measurements of joint angles, despite a complex calibration process prior to each experiment.

Previous studies have also attempted sensor latency compensation. Latency compensation allows for the suppression of time delays due to data processing time in 3D cameras. In our previous study [18], the latency of the Kinect V2 sensor was measured and corrected using the Dynamic Time Warping (DTW) algorithm. The latency compensation process improves the estimation of joint angles during gait analysis.

ANNs have been another method to improve the accuracy of 3D cameras for gait analysis. For example, in the study of Kidziński et al. [19], ANNs were applied to train the detection of foot-contact and toe-off events. Detection of these events is the initial step in the post-processing of most quantitative gait analysis workflows. Also, in the study presented by Bersamira et al. [9], the researchers applied ANNs for the integration of 3D cameras and inertial sensors. The authors claim that the fusion of these systems makes it possible to obtain gait data comparable to those produced by a Vicon system. However, as mentioned above, according to Destelle et al. [17] the use of IMUs together with 3D cameras is an alternative that requires a complex calibration prior to each experiment.

The low accuracy of the measurements obtained using 3D cameras in gait analysis has also been attributed to the joint estimation method. Commonly used Software Development Kits (SDKs) for skeleton tracking estimate the joint center using ML techniques. Firstly, ML is applied to label the pixels corresponding to each body segment. Afterwards, the intersection of these segments is identified to estimate the joint center location [20]. According to Matthew et al. [14], by applying this method, the estimated joint centers may be biologically inconsistent. This can lead to errors in the location of the ankle, knee, and hip, which complicates their use for later gait analysis [1]. This problem encouraged the development of a series of deep learning methods for joint estimation [21,22]. These methods have been particularly applied in self-occlusive tasks such as sitting. Unfortunately, these methods do not study more sophisticated gait descriptors.

In summary, previous studies have attempted to improve the accuracy of the estimation of joint location, but not the gait kinematics itself. The studies that have addressed gait analysis have dealt only with spatio-temporal parameters. However, achieving an accurate enough gait analysis system based on 3D cameras remains a challenge. In contrast to previous studies, this study aims to directly improve the estimation of kinematic gait signals and gait descriptors based on supervised learning through the application of two training approaches.

3. Methodology

3.1. Ethics statement

The current study was approved by the ethics committee of the Faculty of Physical Activity and Sport Sciences INEF-Polytechnic University of Madrid, Spain (protocol code DPI2011-28160-C03-02). The experiments were performed in the Sports Biomechanics Laboratory of INEF. All subjects provided written informed consent prior to participation. In addition, the participants completed an electronic questionnaire with basic questions related to health problems and musculoskeletal injuries suffered in recent years.

3.2. Marker-less motion capture framework and sensor

In the current study, the skeleton tracking task was performed with an Orbbec Astra 3D camera using the 0.34.1 release of the Nuitrack SDK. Both the Orbbec Astra sensor and Nuitrack SDK are widely used and considered powerful for marker-less motion capture and human pose estimation. The following items provide additional details about the system:

- Orbbec Astra: this 3D camera is capable of providing depth sensing with a resolution of 640x480 pixels, 30fps. This allows for accurate depth perception and tracking of human movements. The device uses infrared technology to capture depth information. It employs structured light techniques to calculate depth values for each pixel in the sensor's field of view (H60°, V49.5°, D73°).
- Nuitrack SDK: Nuitrack is a middleware software development kit that offers real-time skeletal tracking and gesture recognition capabilities. It is designed to work seamlessly with various depth sensors, including the Orbbec Astra. This framework employs sophisticated algorithms to analyze the depth information and estimate the positions and orientations of the human body joints, thereby creating a skeleton model.

Certainly, many skeleton tracking frameworks are currently available, however Nuitrack offers several advantages such as easy integration (compatible with Windows, Linux, macOS, Android, and iOS), sensor agnostic (compatibility with Orbbec Astra, Intel RealSense, and Kinect devices) and real-time tracking (results with minimal latency). In addition, one of the most important features is the support for moving sensors, which is essential for the application presented in this paper featuring a mobile robot that analyzes gait. Thanks to this feature the framework can implement skeleton tracking even when the sensor is not static.

3.3. Experimental setup

During the experiments, the 3D camera was mounted on a mobile robot named Rosbot 2.0 from Husarion, as shown in Fig. 1. Using a robot-mounted 3D camera for gait analysis offers several advantages over static 3D cameras. Some of these advantages are listed below:

- Enhanced Coverage: the robot can move autonomously throughout an environment, and can capture gait data from various viewpoints and perspectives, providing a more complete understanding of the person's walking pattern
- Dynamic Gait Analysis: dynamic analysis can provide insights into how gait is influenced by environmental factors, leading to a more accurate assessment.
- Longitudinal Analysis: by following a person over an extended period, a mobile robot can capture gait data longitudinally, allowing for the monitoring of gait changes and adaptations over time.
- Improved Precision: a mobile robot can position itself at optimal distances and angles. This enhanced control over the capture process can result in more precise measurements.
- Flexibility and Adaptability: the mobility of using a robot enables it to be deployed in various settings, such as hospitals, rehabilitation centers, or even home environments.
- Reduction of Observer Bias: the potential for observer bias is minimized. Unlike with static depth sensors, where a human operator must manually position the sensors or make subjective judgments, the robot's automated movement ensures consistency and objectivity in data collection.

Fig. 1. An overview of the experimental setup. Faculty of Physical Activity and Sport Sciences INEF (Polytechnic University of Madrid, Spain).

Table 1
Demographics.

Participants:	37
Gait sequences:	207
Male:	21
Female:	16
Mean age:	21±2 years
Mean weight:	67.41±10.28 kg.
Mean height:	172.99±8.53 cm.

It's important to note that while using a robot-mounted 3D camera provides these advantages, it also introduces challenges such as person following task, and navigation. These factors need to be carefully addressed to ensure accurate and reliable gait analysis results. In this study, the robot is configured to maintain a constant 2.5 meters distance from the person while walking. To achieve this, a controller was implemented in the robot, which was presented in our former studies [23,24]. We encourage the reader to review the aforementioned studies for more information about the robot configuration, navigation and control.

3.4. Data collection

Thirty-seven healthy participants (16 female and 21 male, Table 1) were involved in the study. A total of 207 gait recordings were processed. As mentioned above, skeleton tracking task was performed with an Orbbec Astra 3D camera and the Nuitrack SDK. For each gait sequence, 15 joint trajectories were recorded at 30 Hz. Together with the robot-mounted 3D camera, gait information was retrieved with a Vicon system, which was considered as the ground truth in this study. The Vicon system consisted of six M2 MCAM cameras recording at a sampling rate of 120 Hz. The experiments were conducted in two phases, which are summarized below:

- Participant preparation (15 minutes). Thirty-two reflective markers were placed on anatomical landmarks following the Plug-in-Gait model [25]. The participant wore running shoes, short tights, a top (in the case of women) and socks below the ankles. In addition, the participant signed an informed consent and completed an electronic questionnaire. The height and weight of each participant was registered. Finally, a warm-up and gait speed normalization process was done whereby the subject was prepared to walk at an average speed of 1 m/s.
- Gait data capture (15 minutes). Gait acquisition was performed in a 15x5-meter corridor, in a one-way straight line gait. Six iterations per participant were repeated, giving a total of 222 gait sequences. However, some of these were discarded due to the loss of markers in the Vicon system, whose trajectories could not be reconstructed by filling in gaps. As a result, a total of 207 gait recordings were used for the supervised learning stage.

3.5. Data processing

The robot-mounted 3D camera and the Vicon system started recording data at different times and through different input streams. The timestamp of the first heel strike was used to synchronize the data. This point was retrieved from the anterior-posterior distance between ankle joints (the inter-ankle distance) according to Ceccato et al. [26]. This event represented the onset of the recording. The end of the recording is determined by the last timestamp registered by the Vicon system. The data from both systems was cut between these limits. Due to the difference in sampling frequencies, the Vicon time series were linearly interpolated using the timestamps of the 3D camera data. This process ensures the same query points. Finally, to minimize any fluctuation, a low pass filter with a cut-off frequency of 4 Hz was applied in the 3D camera data.

Table 2
Mapping rules to compare the skeleton model of the 3D camera and the Vicon system.

3D Camera	Vicon Markers
Spine base	Midpoint [RPSI,LPSI,RASI,LASI]
Spine middle	Midpoint [T10,STRN]
Spine shoulder	Midpoint [C7,CLAV]
Left shoulder	Marker LSHO
Left elbow	Midpoint [LELB marker A,LELB marker B]
Left hand	Midpoint [LWR marker A,LWR marker B]
Right shoulder	Marker RSHO
Right elbow	Midpoint [RELB marker A,RELB marker B]
Right hand	Midpoint [RWR marker A,RWR marker B]
Left hip	Hip joint centering algorithm[LASI,LPSI] [a]
Left knee	Midpoint [LKNE marker A,LKNE marker B]
Left ankle	Midpoint [LANK marker A,LANK marker B]
Right hip	Hip joint centering algorithm[RASI,RPSI] [a]
Right knee	Midpoint [RKNE marker A,RKNE marker B]
Right ankle	Midpoint [RANK marker A,RANK marker B]

[a] The hip joint center location was calculated with the hip joint centering algorithm recommended by [25].

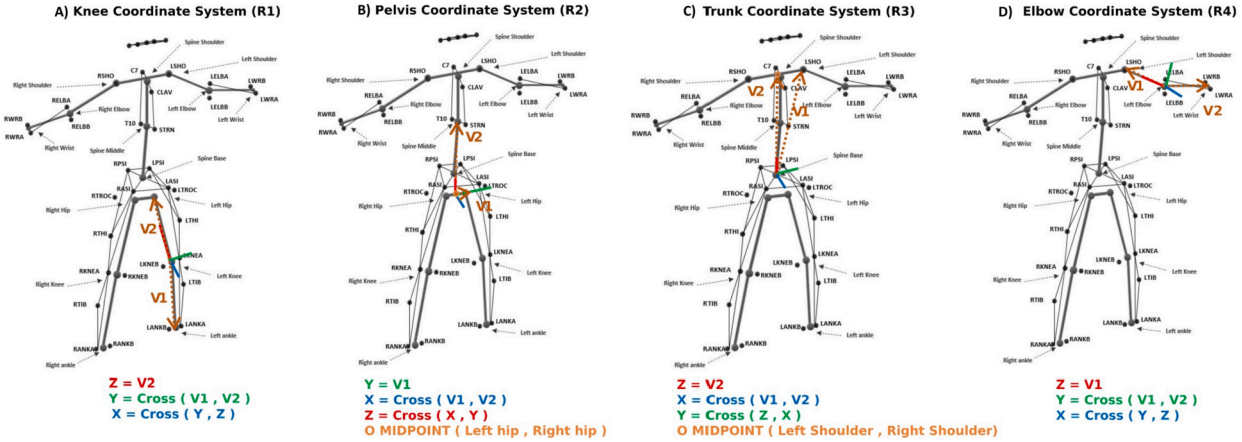

Fig. 2. Coordinate systems definition of the body segments used to apply the inverse kinematics. Local coordinate systems are defined at the A) knee, B) pelvis , C) trunk and D) elbow. For each joint, the coordinate system is indicated with the X-axis in blue, the Y-axis in green, and the Z-axis in red.

3.6. Data analysis

To relate as much as possible the gait model applied by the Nuitrack in the 3D camera with that applied by the Plug-in-Gait in the Vicon system, it was necessary to find an equivalence between the landmarks detected by both systems by defining certain mapping rules. This mapping process is not new. Mapping rules similar to those used in [27,28] were adopted, and are shown in Table 2.

Six Vicon markers (LTROC, LTHI, LTIB, RTROC, RTHI and RTIB) were not considered in this mapping process. For the analysis of gait in this study, the joint angles from elbow, knee, hip, shoulder, trunk, and pelvis were analyzed in sagittal plane. Furthermore, the joint angles from hip, and shoulder were analyzed in frontal plane. In addition, an examination of trunk and pelvis took place in the transverse plane.

3.7. Inverse kinematic model to compute joint angles with the 3D camera

The inverse kinematics is the process to retrieve joint angles from 3D coordinates of landmarks. For the inverse kinematics process in the 3D camera, the coordinate systems definition of the body segments were compatible with the International Society of Biomechanics recommendations [29]. First, four local coordinate systems are defined at the joints: $R1 = {}^{w}F_{Knee}$, $R2 = {}^{w}F_{Pelvis}$, $R3 = {}^{w}F_{Trunk}$, $R4 = {}^{w}F_{Elbow}$. These are the projected knee, pelvis, trunk, and elbow coordinate frames, on the world frame, and are defined as shown in Fig. 2.

To locate these coordinate systems, vectors V1 and V2 are defined in alignment with the body segments involved. In Fig. 2, the coordinate axes are color-coded: X in blue, Y in green and Z in red. In this figure, at the bottom of each skeleton, the definition of each coordinate axis is indicated according to the order in which they are defined. For example, in the case of the knee coordinate system $R1$, first the Z axis is defined as parallel to the vector V2, which connects knee to hip. Then the Y-axis is defined as the cross product between V1 and V2, where V1 is the vector connecting knee to ankle. Finally, X is defined as the cross product between Y

and Z. In some cases, in order to define the coordinate system, it was necessary to calculate a virtual landmark at the midpoint of two adjacent landmarks. These midpoints are colored in orange. For example, the midpoint of left and right hip was used to define the coordinate system $R2$ and the midpoint between left and right shoulder was used to define the coordinate system $R3$.

Once the local coordinate systems are defined, the mathematical definitions of the joint angles correspond as closely as possible to the existing clinical terminology as follows:

- The hip joint angles were determined as the Cardan rotation angles in the sequence YXZ between the pelvic and knee coordinate systems, i.e. the relative rotation $Eyxz1$, where $Eyxz1 = R1 \cdot R2^T$.
- The pelvic angles were calculated using the Cardan rotation angles in the sequence YXZ between the world and pelvis coordinate frames. This corresponds to the absolute rotation of $R2$ with respect to the world frame.
- The Cardan rotation angles between the world coordinate and the trunk coordinate frames were used to determine the trunk angles. This corresponds to the absolute rotation of $R3$ with respect to the world frame.
- The angle between the two vectors crossing at the knee joint center was considered as the knee joint angle. These vectors were rotated according to $R1$.
- The angle between the two vectors crossing at the elbow joint center was considered as the elbow joint angle. These vectors were rotated according to $R4$.
- The shoulder joint angles were determined as the Cardan rotation angles in the sequence YXZ between the trunk and elbow coordinate systems, i.e. the relative rotation $Eyxz2$, where $Eyxz2 = R4 \cdot R3^T$.

4. Learning process: training, validation and testing

The large number of experiments performed together with the robot-mounted 3D camera and the Vicon system allowed the application of supervised learning algorithms to improve the resulting estimations. This application was only possible due to the nature of our study in which the signals can be post-processed before obtaining the final results. There are two basic possible approaches for training and both have been studied in order to see which one achieves a better result: training based on kinematic gait signals, and training based on gait descriptors.

As mentioned above, 37 participants provided 207 gait recordings. Of these, the recordings corresponding to 20% of the participants were excluded from the learning process (i.e. were not used for training, validation or testing). This represents the re-testing group of participants. The data from these participants was used to make a further comparison between both training approaches.

Given that the accuracy of Orbbec Astra is quite different for signals in the sagittal plane than for signals in the frontal and transverse planes, combining the signals in a multi-channel neural network could reduce accuracy and lead to less effective results. Therefore, in the approach presented in this paper, independent neural networks were trained for each analyzed signal. This approach, known as multi-task learning, allows each network to focus on a specific signal and optimize its performance independently.

4.1. Training approach based on joint kinematics

Since the samples from the Vicon system were linearly interpolated using the timestamps of the 3D camera, the samples from both systems are temporally aligned. This allowed us to associate each sample of the camera with a corresponding label on the Vicon signal.

Each kinematic signal was represented by a single vector of dimension Nx1, where N was the total number of samples collected from all the experiments. From this single vector, a training matrix was generated by applying a windowing process. The windowing process is used to split the signal into several frames, or windows. The skip factor between frames was set to one sample. Therefore, the final training matrix had a dimension of (N-W+1)xW, where W was the windowing width. On this basis, each training sample was a vector of 1xW that had a corresponding target of dimension 1x1 coming from the Vicon signal.

This data set was randomly split into three classes: training, validation and testing. The partitioning followed the rule 60/20/20, as follows:

- 60% of the data was used for training.
- 20% of the data was used to validate the network and stop training before overfitting occurs.
- 20% of data was used as a completely independent test of network generalization.

A two-layer feedforward backpropagation ANN was created and trained per each kinematic signal using multi-task learning. The ANN architecture had 10 neurons in the hidden layer, an input layer with a number of entries equal to the windowing width factor and an output layer with a single outcome. The transfer (activation) function for the first layer was a Sigmoid Function and for the second layer was a Linear Function. Experimentally, a windowing width of 3 samples was determined to be the most effective. The training algorithm updated weight and bias values according to Levenberg-Marquardt optimization. Mean squared error (MSE) was used as the performance function. The maximum number of training iterations was set to 1000 epochs. More details of the specific training parameters are given in Table 3 and a schematic of the chosen ANN architecture is shown in Fig. 3.

Table 4 summarizes the accuracy of the system before and after training. When analyzing the correlations shown in Table 4, it can be seen that before training, trunk tilt, pelvis tilt, hip add./abd., shoulder add./abd. and pelvis rotation had correlations lower than 0.70. The low accuracy in shoulder and hip add./abd. may be attributed to the principle operation of 3D cameras. Since

Table 3
ANN configuration and training parameters.

Data Configuration for NN	Training Parameters
Data Division: Random	Epochs (max): 1000
Training: Levenberg-Marquardt	Time: Inf
Performance function: MSE	Goal: 0
Predictors: [3x7140]	Min. grad: 1e-07
Responses: [1x7140]	Max. fail: 6
Training data: 60%	Mu: 0.001
Validation data: 20%	Mu. dec: 0.1
Test data: 20%	Mu. inc: 10

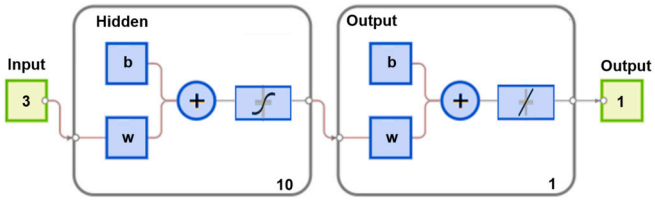

Fig. 3. The standard feedforward backpropagation ANN with 2 hidden layers, 10 neurons in the first layer and 1 neuron in the second layer used to train on each kinematic signal.

Table 4
Accuracy of the estimation provided by the 3D camera in sagittal, frontal and transverse planes before and after training. Results are evaluated using the root-Mean-Square Error (RMSE), the Pearson Correlation (r) and the Mean Absolute Error (MAE). Values are displayed in the format mean±SD.

	BEFORE TRAINING			AFTER TRAINING		
	RMSE	r	MAE	RMSE	r	MAE
Elbow flex/ext. (°)	3.54±0.18	0.79±0.02	2.67±0.06	2.03±0.08	0.8±0.02	1.57±0.11
Knee flex/ext. (°)	6.89±0.29	0.94±0.01	5.25±0.19	5.94±0.31	0.95±0.01	4.39±0.21
Hip flex/ext. (°)	11.38±0.24	0.93±0.01	10.07±0.13	4.52±0.19	0.95±0	3.53±0.25
Shoulder flex/ext. (°)	10.03±0.28	0.69±0.03	8.37±0.18	6.48±0.4	0.71±0.03	5.04±0.24
Trunk tilt (°)	5.39±0.11	0.45±0.04	4.76±0.06	1.94±0.06	0.55±0.03	1.54±0.11
Pelvis tilt (°)	5.81±0.11	0.43±0.04	5.27±0.05	1.91±0.06	0.55±0.03	1.54±0.12
Hip add/abd. (°)	4.79±0.11	0.64±0.03	4.11±0.07	2.59±0.1	0.7±0.03	2.04±0.12
Shoulder add/abd. (°)	7.65±0.22	0.68±0.03	6.69±0.11	3.27±0.25	0.73±0.03	2.36±0.17
Trunk rotation (°)	2.77±0.15	0.82±0.02	2±0.05	1.72±0.11	0.87±0.02	1.23±0.08
Pelvis rotation (°)	6.95±0.18	0.62±0.03	5.8±0.07	2.71±0.09	0.71±0.02	2.19±0.17

the 3D camera does not provide useful information for the detection of motion in frontal plane, the SDK relies on the intrinsic parameters of the RGB camera for the estimation, so these motions are approximations of medium quality. However, after training, the improvements are remarkable. Just trunk tilt and pelvis tilt remain with correlation lower than 0.70. Despite this fact, considering the low range of motion of these joints (about 2 degrees), the results are satisfactory.

On the other hand, joint kinematics could be affected by the difference in the gait models between Vicon and the 3D camera. Since the number of markers detected by the camera is smaller than those detected by the Vicon, the definition of the coordinate systems of the body segments is in fact different in each system. While the Vicon uses four markers to detect the kinematics of the pelvis, the 3D camera uses only two landmarks, plus one belonging to the middle spine. However, the improvement in the correlation of the signals shown in Table 4, reflects a fit of the models. Thus, the kinematic signals reported by the 3D camera more closely resembles that of the Vicon.

Fig. 4 shows the effects of the training process for kinematic signals of gait in the re-testing group of participants. From the observation of Fig. 4, another problem arising from the difference in gait models can be noted. In some of the signals, such as the hip flex./ext., shoulder flex./ext., trunk tilt, pelvis tilt and shoulder add./abd., there is an offset with respect to the signal of the Vicon system. The correction of this offset is another benefit of the learning process that provides a better interpretation of joint kinematics. In the case of the trunk and pelvis tilt, this offset could not be corrected completely. The randomness of this offset in the training data and the small range of motion of these joints (about 2 degrees) made it difficult for the ANN to correct it. However, a considerable improvement in the correlation of these signals was obtained.

It is very important to emphasize that training is not intended to learn only the normal gait pattern performed by the participants. This would make the system unsuitable for the analysis of new gait patterns or pathological patterns. On the contrary, what the ANN actually learns is how to correct the measurement errors of the original gait model in the 3D camera, based on the examples provided by the Vicon system. To achieve this, the ANN does not receive as input the entire gait cycle in each sample. Instead, it receives the

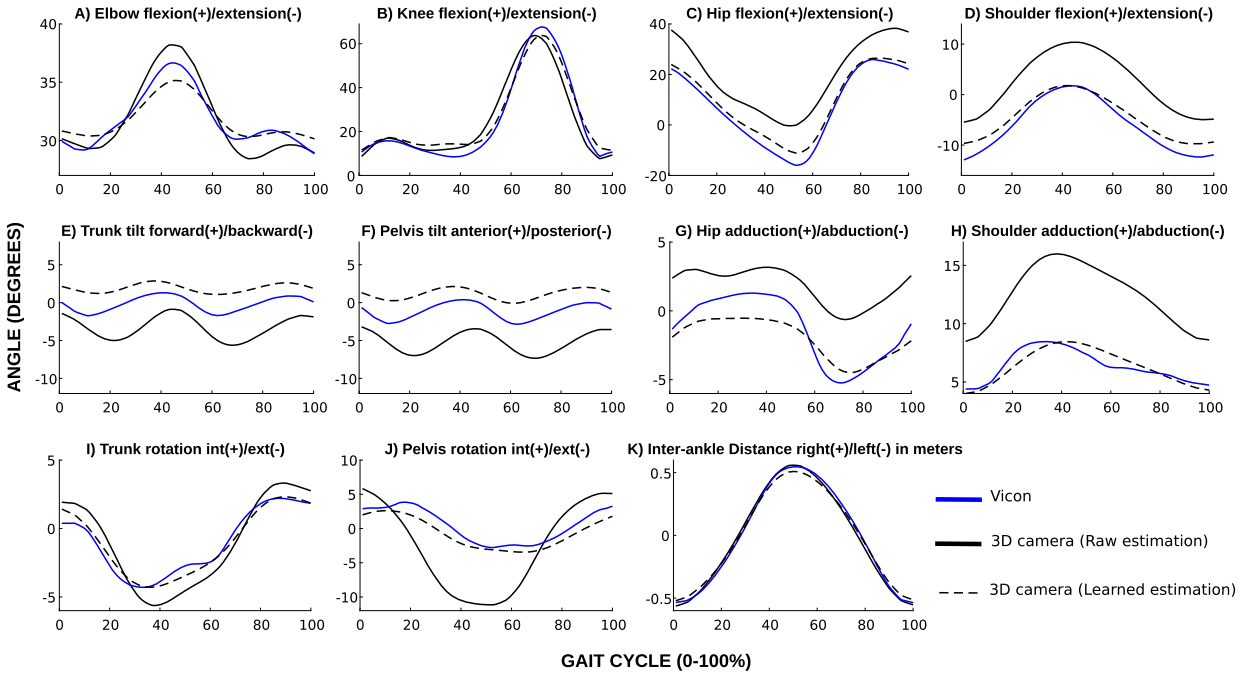

Fig. 4. Kinematic gait cycles retrieved from the re-testing group of participants. The gait cycles are normalized from 0-100% and averaged for all the iterations. The figure shows the Vicon system (blue line), the raw estimation of 3D camera (continuous black line), and the learned estimation of 3D camera (dashed black line). The panels are ordered as follows: signals from A to F belong to the sagittal body plane, those from G to H to the frontal body plane and those from I to K to the transverse body plane.

information into data frames thanks to the windowing process. This is extremely important for generalizing learning. Consequently, it is possible to apply the training result to the analysis of new gait patterns.

4.2. Training approach based on gait descriptors

This second approach arises from the concern of whether it would be better to train the ANN directly with gait descriptors rather than with kinematic signals divided into data frames. This is a more direct way to teach the ANN to report gait descriptors more accurately.

In order to apply this second approach 27 most important descriptors representing the main differentiators of normal and pathological gait were selected for the analysis following the recommendation of Molina and Carratalá [8]. As listed in Table 5, 11 were spatio-temporal descriptors and 16 were kinematic descriptors.

Fig. 5 illustrates the process followed for the identification of the set of kinematic gait descriptors. According to Ceccato et al. [26], the reference signal for the detection of heel strike and toe-off events was the maximal anterior-posterior distance between ankle joints (or inter-ankle distance). In this signal there is a characteristic sinusoidal curve when the X coordinate of the foot markers is graphed versus time. In this curve, the valleys correlate to the time at which the foot comes into contact with the ground (the heel strike) and the peaks fit with the initiation of swing phase (the toe-off) [30]. Fig. 5 illustrates the process followed for the identification of kinematic descriptors of knee and hip in sagittal plane. The rest of the kinematic descriptors correspond to instants of maxima and minima in the corresponding kinematic signals. In addition, the set of spatio-temporal descriptors were calculated according to established definitions [12,31,32].

Each gait descriptor was characterized by a single vector of dimension Px1, with P being the number of gait cycles collected from all the experiments. This means that a unique sample is extracted from each gait cycle. Given that 27 descriptors were analyzed, a training matrix of dimension Px27 was generated. A target matrix of the same dimensions was obtained with the estimations of the Vicon system. Corresponding columns were used to train independent neural networks applying a multi-task learning approach.

Each ANN used a two-layer feedforward backpropagation architecture with 10 neurons in the hidden layer, an input layer with a single entry, and an output layer with a single outcome (same as shown in Fig. 3). The training algorithm updated weight and bias values according to Levenberg-Marquardt optimization, and MSE was used as the performance function. The maximum number of training iterations was set to 1000 epochs. Again the data set was split into three classes: training, validation and testing, following the rule 60/20/20 (60% for training, 20% for validation and 20% for testing).

Table 5 summarizes the results of this training approach. Results show that the spatio-temporal parameters have no further improvement after the training process. An exception occurs in the case of left/right stride length, where a noticeable correction is observed. Regarding the sixteen kinematic descriptors, the results are always better after training. Especially the descriptors for hip, pelvis and trunk show remarkable improvements.

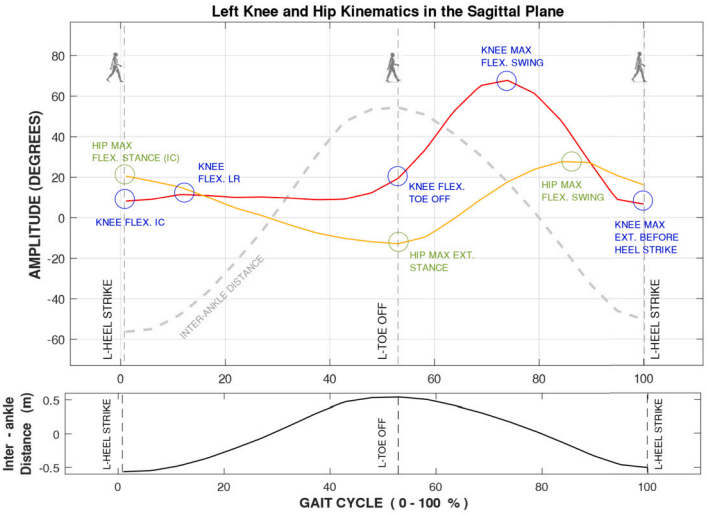

Fig. 5. Identification of kinematic descriptors for left knee and hip in sagittal plane. The reference signal for the detection of heel strike and toe-off events was the inter-ankle distance [26].

Table 5
Accuracy of the system for the detection of 11 spatio-temporal and 16 kinematic descriptors. The table shows the results before and after training. Results are evaluated using the Root-Mean-Square Error (RMSE) and the Mean Absolute Error (MAE). Values are displayed in the format mean±SD.

GAIT DESCRIPTORS	BEFORE TRAINING		AFTER TRAINING	
	RMSE	MAE	RMSE	MAE
SPATIO-TEMPORAL DESCRIPTORS				
step width (m)	0.022±0.01	0.01±0.01	0.021±0.01	0.02±0.01
left stride length (m)	0.15±0.02	0.14±0.03	0.07±0.01	0.05±0.02
right stride length (m)	0.13±0.03	0.11±0.03	0.05±0.01	0.04±0.01
left stride time (s)	0.03±0.01	0.02±0.01	0.04±0.01	0.03±0.01
right stride time (s)	0.06±0.02	0.04±0.02	0.06±0.02	0.04±0.01
right step time (s)	0.04±0.02	0.02±0.01	0.03±0.01	0.02±0.00
left step time (s)	0.027±0.01	0.02±0.01	0.026±0.01	0.02±0.01
right cadence ($steps/min$)	4.65±1.08	3.29±1.4	4.60±1.27	3.39±1.36
left cadence ($steps/min$)	3.60±1.80	2.05±1.37	4.01±1.51	2.99±1.15
percentage of foot stance (%)	3.74±0.78	3.00±1.12	2.74±0.73	2.33±0.59
percentage of foot swing (%)	2.73±1.41	1.78±1.03	1.32±0.29	1.08±0.19
KINEMATIC DESCRIPTORS				
trunk max. tilt (°)	4.22±0.87	3.59±0.63	1.68±0.45	1.26±0.23
trunk min. tilt (°)	5.93±1.21	5.41±1.33	2.18±0.35	1.92±0.29
pelvis max. tilt (°)	4.31±0.57	3.87±0.56	1.33±0.52	1±0.39
pelvis min. tilt (°)	6.14±0.64	5.91±0.75	1.67±0.31	1.38±0.31
hip max. adduction (°)	3.86±0.75	3.49±0.67	1.37±0.35	1.09±0.38
hip min. abduction (°)	5.02±0.54	4.64±0.67	2.22±0.67	1.66±0.47
pelvis max. rotation (°)	6.72±1.59	5.99±1.43	2.96±0.97	2.5±0.53
pelvis min. rotation (°)	9.81±0.98	9.27±1.35	2.02±0.47	1.63±0.39
hip max. extension during stance (°)	14.08±1.53	13.44±1.46	3.6±0.83	2.96±0.87
hip max. flexion during swing (°)	12.22±1.31	11.75±1.88	2.61±0.54	2.37±0.33
hip max. flexion during stance (°)	15.10±1.47	14.59±1.35	3.7±1.05	3.23±0.44
knee initial contact position (°)	7.43±1.97	5.99±1.77	4.91±1.33	3.71±1.28
knee position at toe-off (°)	9.85±1.81	8.77±2.15	3.29±1.03	2.44±0.66
knee max. flexion in load response (°)	4.10±1.02	2.93±0.90	4.52±1.03	3.90±0.90
knee max. flexion during swing (°)	4.06±1.21	3.04±0.61	3.39±1.69	2.37±0.98
knee max. extension before heel strike (°)	5.76±1.68	4.53±1.75	4.81±1.52	3.57±0.95

5. Comparison of training approaches

This section is focused on the comparison of both training approaches. Specifically, it compares the accuracy for the estimation of the sixteen kinematic descriptors of gait. For this purpose, the re-test group of participants was used. Fig. 6 shows the process followed to obtain the data for comparison. On the basis of this process, four estimations can be compared: the Vicon estimation (V); the 3D camera raw estimation (C); the 3D camera trained with kinematic signals (CTs); and the 3D camera trained with gait descriptors (CTd).

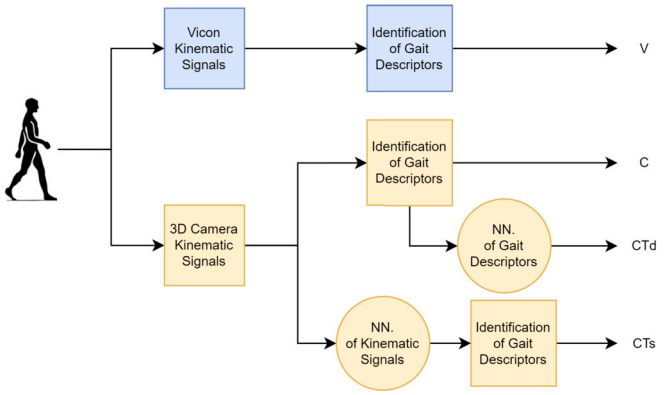

Fig. 6. Process followed to obtain the data for the comparison of the two training approaches.

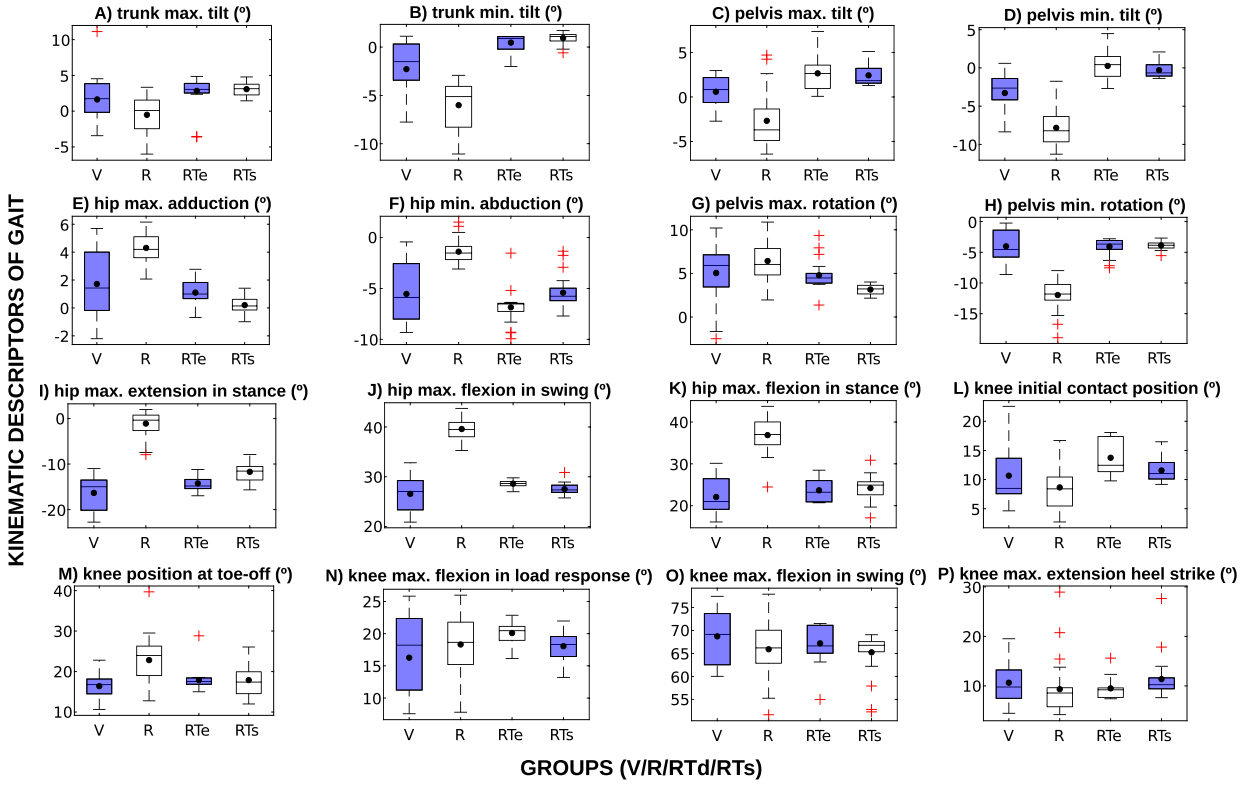

Fig. 7. Results of the comparison of both training approaches. The box plots show the 25th percentile, 75th percentile, median, minimum and maximum values. The mean value is represented with a black dot. The box plot closest to the Vicon is colored in blue, which is also colored in blue. The panels describe results for the following gait events: A-B trunk flexion events, C-D pelvis flexion events, E-F hip adduction and abduction events, G-H hip rotation events, I-K hip flexion events, L-P knee flexion events.

Results of this comparison are shown in the Fig. 7. A visual representation of the data is shown using box plots. These box plots summarize the 25th percentile, 75th percentile, median, minimum and maximum values of each group of data. The mean value is also shown with a black dot over the box plots. In this figure, the box plot belonging to the system with mean value closest to that of the Vicon is colored in blue, which is also colored in blue.

When analyzing the sixteen kinematic descriptors shown in Fig. 7, it is clear that the trained systems (CTd or CTs) are always better than the untrained system (C). When comparing the two training approaches, it is observed that in 62% of the events analyzed (10/16) the approach based on gait descriptors (CTd) is better than the approach based on kinematic signals (CTs), where only in 38% of the cases (6/16) it proved to be better. However, it can be observed that for events of large amplitude, such as maximum knee and hip flexions and extensions, training based on gait descriptors (CTd) is better. On the contrary, for kinematics of smaller range of motion, the training based on kinematic signals (CTs) is superior. This can be noted in the pelvis maximum and minimum tilt, and in the lower amplitude descriptors of knee kinematics such as the knee initial contact position, knee maximum flexion in load response,

Table 6
Comparison of our study with current approaches.

	OUR SYSTEM[a]		RIGID BODIES[b]		DTW[c]		MULTI-SENSOR[d]		MULTI-SENSOR[e]		3D CAMERA+IMU[f]		3D CAMERA+IMU[g]	
	ROOT MEAN SQUARE ERROR													
	BEF.	AFT.	BEF.	AFT.	BEF.	AFT.	BEF.	AFT.	BEF.	AFT.	BEF.	AFT.	BEF.	AFT.
Elbow flex./ext.	3.54	2.03	-	-	-	-	-	-	-	-	-	5.90	13.33	9.00
Knee flex./ext.	6.89	5.94	9.40	5.20	13.15	10.47	-	-	-	-	-	-	28.23	7.89
Hip flex./ext	11.38	4.52	11.10	5.60	6.20	2.65	-	-	-	-	-	-	-	-
Shoulder flex./ext.	10.03	6.48	-	-	9.53	5.80	-	-	-	-	-	4.70	-	-
Trunk forw./back. tilt	5.39	1.94	3.90	3.50	2.09	1.93	-	-	-	-	-	-	-	-
Pelvis ant./post. tilt	5.81	1.91	-	6.10	3.22	3.14	-	-	-	-	-	-	-	-
Hip add./abd.	4.79	2.59	-	-	5.83	4.70	-	-	-	-	-	-	-	-
Shoulder add./abd.	7.65	3.27	-	-	11.76	11.69	-	-	-	-	-	-	-	-
Trunk int./ext. rot.	2.77	1.72	-	-	5.26	5.09	-	-	-	-	-	-	-	-
Pelvis int./ext. rot.	6.95	2.71	-	-	5.88	4.81	-	-	-	-	-	-	-	-
step width (m)	0.022	0.021	-	-	-	-	-	0.01	0.0066	0.0039	-	-	-	-
AV. stride length (m)	0.14	0.06	-	-	-	-	-	0.00	0.0004	0.0020	-	-	-	-
AV. stride time (s)	0.05	0.05	-	-	-	-	-	0.01	-	-	-	-	-	-
AV.step time (s)	0.034	0.028	-	-	-	-	-	0.01	0.000	0.000	-	-	-	-
AV. Cadence (steps/min)	4.12	4.31	-	-	-	-	-	0.90	-	-				

[a] Our trained robot-mounted 3D camera.
[b] Matthew et al. [14]. Single Kinect, tested in a sit to stand task. The original joint data is transformed into a new model built with rigid bodies.
[c] Guffanti et al. [18]. Correction of latency usind DTW(Dynamic Time Warping).
[d] Geerse et al. [11]. Multi-Kinect v2, 10-Meter Walkway for spatio-temporal gait analysis.
[e] Müller et al. [12]. Multi-Kinect v2, 7-Meter Walkway for spatio-temporal gait analysis.
[f] Alizadegan and Behzadipour [16]. Single Kinect to improve accuracy of inertial joint angle estimation in upper limb stroke rehabilitation.
[g] Destelle et al. [17]. Single Kinect in front of the person and fusion with IMUs. The RMSE values were averaged for the two experiments performed by the authors.

and in the knee maximum extension before heel strike. Therefore, it is not possible to define which training approach is completely better. Despite the fact that training based on gait descriptors (CTd) presented higher accuracy in most parameters, training based on kinematic signals (CTs) has an enormous advantage because it allows to know the complete kinematics of the joint, which is very useful when determining deviations between gait patterns for clinical purposes. Consequently, the selection of the training approach will depend on the purpose of the study conducted.

6. Discussion

It is interesting to compare the results obtained with other techniques for improving the estimations of 3D cameras. An in-depth comparison can be made by analyzing Table 6. Current approaches have attempted to improve the accuracy of 3D cameras in gait analysis through remodeling the skeleton, sensor fusion, or applying machine learning techniques.

In the study presented by Bersamira et al. [9], the researchers claim to have developed an integrated system consisting of a 3D camera and IMUs comparable to the quality produced by a Vicon system. Bayesian Regularization ANN (BRANN) was used to integrate both sensors using Vicon data as the target. Bayesian regularization is one of the regularization methods used to compensate for ANN problems such as slow convergence and overfitting of training patterns. Although this study is interesting, mainly because of the number of participants, major concerns arise when analyzing the methodology followed. Nothing is stated about the definition for determining the spatio-temporal and angular parameters. Also it is not possible to know how the training was generalized to prevent the algorithm from learning a specific gait pattern. On the contrary, in our study the number of participants is also high but caution is taken that the ANNs do not learn a specific gait pattern thanks to the windowing process applied.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are two popular types of Recurrent Neural Networks (RNNs) that have been shown to be effective for modeling sequential data (time series). LSTM and GRU architectures have been used to improve gait signals captured by RGBD sensors in various studies. For instance, the study presented by Challa et al. [33] shows promising results. The errors of the predicted signals are low, although the analyzed gait is on a treadmill and the camera (a Microsoft Kinect) is static. Similarly, L. Konz et al. [34] used a combination of CNNs and LSTM networks for human gait recognition from RGBD and depth data captured by a Kinect sensor. The study also showed that the LSTM network was effective in capturing the temporal dependencies in the gait signals. Luu et al. used Generalized Regression Neural Networks (GRNN) and inverse fast Fourier transforms to predict joint angles of lower limbs. Results showed small errors, in the range of 3-5 degrees, compared to others, but the test data were included in the training data set, which would produce less-accurate predictions when validation is performed.

Overall, LSTM and GRU architectures have shown promise for improving the gait signals captured by RGBD sensors. However, other studies such as the presented by Lim et al. [35] have preferred (as in our case) the use of simple ANNs to improve gait signals. In the study of Lim et al., data from an IMU is fed into an ANN, which consists of a single hidden layer of 20 activated sigmoidal neurons and eleven output neurons. However, unlike our study, that study uses signals from an IMU sensor instead of a depth sensor.

In the study presented by Müller et al. [12], a validation of enhanced Kinect sensor for gait assessment was presented. The idea of that study was to compute a spatially averaged skeleton tracked by two Kinect sensors placed at both sides of the walking path. The study demonstrated that sensor setups tracking the person only from one-side is less accurate and should be replaced by two-sided setups. To fuse data from both sensors, the authors used the iterative closest point method, which adds imprecision to the model. Similarly, the use of more than one camera suggests a requirement of a dedicated room where the system can be setup and left undisturbed between sessions.

Regarding the fusion of 3D cameras with inertial sensors (IMUs), it is interesting to mention the study presented by Destelle et al. [17]. The authors built a fused Kinect/IMUs skeleton where the Kinect sensor provided the positions of the joints and the IMUs provided the orientation information. The use of IMUs together with 3D cameras required a complex calibration prior to each experiment. The authors reported a final accuracy of 9.00° for elbow flex./ext., and 7.89° for knee flex./ext. In contrast, in our study considerably higher accuracies were achieved after training (2.03° for elbow flex./ext. and 5.94° for knee flex./ext.). It should also be noted that the study of Destelle et al. [17] was performed for a simulated kicking motion while in our study a real gait was analyzed.

In the study presented by Alizadegan and Behzadipour [16], the authors proposed a new method to improve accuracy of IMUs for joint angle estimation in upper limb rehabilitation applications. To correct for the sensor-to-segment misalignment of the inertial sensors, position measurements retrieved from Kinect sensor were used. The experiment was performed on a mechanical upper limb stroke rehabilitation device also known as shoulder wheel. The shaft rotation angle was encoded and this measurement was used as ground truth. The authors reported accuracies of 5.90° for elbow flex./ext. and 4.70° for shoulder flex./ext. Conversely, in our study, the elbow is reported with higher accuracy (3.54° for elbow flex./ext.) but the shoulder is slightly less accurate (6.48° for shoulder flex./ext.). It should also be noted that the experimental basis in the study of Alizadegan and Behzadipour [16] is different from ours. While in that study the person remains static, in our study the person is in constant walking motion.

Matthew et al. [14] proposed the improvement of the estimations provided by 3D cameras by remodeling the skeleton to a new one based on rigid bodies. The idea of the authors was to rescale the skeleton detected by the 3D camera, using the height of the person. Then, the joint centers are recalculated based on anthropometric approximations. The method was tested only for a sit to stand task (STS), and the results obtained were satisfactory. However, it should be noted that, applying joint center relocation based on anthropometric approximations may not be a robust technique. This is because 3D cameras are affected by many other factors like light conditions, sensor position, occlusion, among others [4]. This makes it difficult to relocate joint centers using only anthropometric approximations. Thus, this method may not be effective in all experimental conditions. In addition, our method based on ML obtained better results mainly for trunk and pelvis tilt. While in the study of Matthew et al. [14] errors of 3.50° for trunk tilt and 6.10° for pelvis tilt were obtained, in our study 1.94° and 1.91° were achieved respectively.

Finally, in our previous study [18], latency correction was proposed using Dynamic Time Warping (DTW). Latency of 3D cameras produces non-symmetrical temporal variations in signals that directly affects the system performance. Results showed a noticeable improvement in system performance. However, these corrections only solve the problem of the temporal disturbances produced by the variable latency of the sensor. Instead, the application of machine learning not only corrects these distortions. This can be observed in the improvement of the correlation in Table 4 and in the correction of waveforms in Fig. 4. In addition to the above, machine learning can also correct discrepancies in the gait model applied by the 3D camera with respect to the Plug-in-gait model applied by the Vicon system. This effect can be observed in Fig. 4, where the offsets with respect to the Vicon signals were corrected.

7. Conclusion

In this study, it was possible to improve the accuracy of 3D cameras for human gait analysis applications. This was achieved by post-processing the raw estimations of the 3D camera using ANNs trained with the data provided by a certified Vicon system. The neural networks were trained based on two different criteria: training based on kinematic gait signals, and training based on gait descriptors. In the latter approach, 11 spatio-temporal and 16 kinematic descriptors representing the main identifiers of normal and pathological gait were selected for the analysis.

The accuracy of the estimations with the 3D camera was measured before and after training. Results showed lower errors and higher correlations with respect to the gold standard, thus, improving the obtained biomechanical model with the 3D camera. The accuracy obtained when detecting the main descriptors of a pathological gait also showed a substantial improvement, mainly for kinematic descriptors.

A multi-task learning approach was applied, from which independent ANNs were trained for each signal analyzed. This allowed each network to focus on a specific signal and optimize its performance independently. However, a multi-input neural network may be a promising approach to improve multiple gait signals simultaneously, and this would be something to address as potential future work.

When comparing both training approaches, it was not possible to define which was the better in absolute terms. Regardless, training based on kinematic signals has an enormous advantage because it allows to know the complete kinematics of the joint. This is very useful when determining deviations between gait patterns along a long-term clinical analysis. Therefore, we believe that the selection of the training approach will depend on the purpose of the study to be conducted.

This study reveals the great potential of 3D cameras and encourages the research community to continue exploring their use in gait analysis.

Ethics statement

All participants gave their informed consent for inclusion before they participated in the study. The current exploratory study was approved by the Ethics Committee of the Universidad Politécnica de Madrid (DPI2011-28160-C03-02), Spain.

Author contribution statement

Diego Guffanti: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. Alberto Brunete; Miguel Hernando; David Álvarez: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper. Javier Rueda; Enrique Navarro: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The research leading to these results has received funding from PID2020-118299RB-I00 - SISTEMA ROBOTICO NO INVASIVO PARA EL ANALISIS BIOMECANICO DE LA MARCHA HUMANA -, “Convocatoria 2020 Proyectos de I+D+i - RTI Tipo B”, of the Government of Spain. The authors would like to thank Faculty of Physical Activity and Sports Sciences - INEF, UPM, for the use of Sports Biomechanics Laboratory.

References

[1] A. Pfister, A.M. West, S. Bronner, J.A. Noah, Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis, J. Med. Eng. Technol. 38 (5) (2014) 274–280.
[2] B.F. Mentiplay, L.G. Perraton, K.J. Bower, Y.-H. Pua, R. Mcgaw, S. Heywood, R.A. Clark, Gait assessment using the Microsoft Xbox one kinect: concurrent validity and inter-day reliability of spatiotemporal and kinematic variables, J. Biomech. 48 (10) (2015) 2166–2170.
[3] S. Springer, G. Seligmann, Validity of the Kinect for gait assessment: a focused review, Sensors 16 (2), https://doi.org/10.3390/s16020194.
[4] W. Lemkens, P. Kaur, K. Buys, P. Slaets, T. Tuytelaars, J. Schutter, Multi RGB-D camera setup for generating large 3D point clouds, in: IEEE Int. Conf. Intell. Robot. Syst., 2013, pp. 1092–1099.
[5] S. Shrivastava, J. Bharti, R.K. Pateriya, Machine learning based gait abnormality detection using Microsoft kinect sensor, Mater. Today Proc. (2021), https://doi.org/10.1016/j.matpr.2020.11.615, https://www.sciencedirect.com/science/article/pii/S2214785320392890.
[6] H. Zhen, M. Deng, P. Lin, C. Wang, Human gait recognition based on deterministic learning and kinect sensor, in: 2018 Chinese Control Decis. Conf., 2018, pp. 1842–1847.
[7] M. Li, S. Tian, L. Sun, X. Chen, Gait analysis for post-stroke hemiparetic patient by multi-features fusion method, Sensors (Switzerland) 19 (7), https://doi.org/10.3390/s19071737.
[8] M. Francisco, M. Carratalá, La Marcha Humana: Biomecánica, Evaluación y Patología, Editorial Médica Panamericana, Madrid, 2020, https://www.medicapanamericana.com/es/libro/La-marcha-humana-incluye-version-digital.
[9] J.N. Bersamira, R.J.A. De Chavez, D.D.S. Salgado, M.M.C. Sumilang, E.R. Valles, E.A. Roxas, A.R. dela Cruz, Human gait kinematic estimation based on joint data acquisition and analysis from IMU and depth-sensing camera, in: 2019 IEEE 11th Int. Conf. Humanoid, Nanotechnology, Inf. Technol. Commun. Control. Environ. Manag. (HNICEM), 2019, pp. 1–6.
[10] K.-Y. Yeung, T.H. Kwok, C. Wang, Improved skeleton tracking by duplex kinects: a practical approach for real-time applications, J. Comput. Inf. Sci. Eng. 13 (2013) 41007, https://doi.org/10.1115/1.4025404.
[11] D. Geerse, B. Coolen, M. Roerdink, Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments, PLoS ONE 10 (10), https://doi.org/10.1371/journal.pone.0139913.
[12] B. Müller, W. Ilg, M.A. Giese, N. Ludolph, Validation of enhanced kinect sensor based motion capturing for gait assessment, PLoS ONE 12 (4) (2017) 1–18, https://doi.org/10.1371/journal.pone.0175813.
[13] A. Amini, K. Banitsas, An improved technique for increasing the accuracy of joint-to-ground distance tracking in kinect v2 for foot-off and foot contact detection, J. Med. Eng. Technol. 43 (1) (2019) 8–18.
[14] R.P. Matthew, S. Seko, R. Bajcsy, J. Lotz, Kinematic and kinetic validation of an improved depth camera motion assessment system using rigid bodies, IEEE J. Biomed. Health Inform. 23 (4) (2019) 1784–1793.
[15] J. Nichols, M. Sena, J. Hu, O. O’Reilly, B. Feeley, J. Lotz, A kinect-based movement assessment system: marker position comparison to vicon, Comput. Methods Biomech. Biomed. Eng. (2016) 1–11, https://doi.org/10.1080/10255842.2017.1340464.
[16] A. Alizadegan, S. Behzadipour, Shoulder and elbow joint angle estimation for upper limb rehabilitation tasks using low-cost inertial and optical sensors, J. Mech. Med. Biol. 17 (2) (2017) 1750031.
[17] F. Destelle, A. Ahmadi, N.E. O’Connor, K. Moran, A. Chatzitofis, D. Zarpalas, P. Daras, Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors, in: 2014 22nd Eur. Signal Process. Conf., EURASIP, 2014, pp. 371–375.
[18] D. Guffanti, A. Brunete, M. Hernando, J. Rueda, E. Navarro Cabello, The accuracy of the Microsoft kinect V2 sensor for human gait analysis. A different approach for comparison with the ground truth, Sensors 20 (16) (2020) 4405, https://doi.org/10.3390/s20164405, https://www.mdpi.com/1424-8220/20/16/4405.
[19] Ł. Kidziński, S. Delp, M. Schwartz, Automatic real-time gait event detection in children using deep neural networks, PLoS ONE 14 (1) (2019) 1–11, https://doi.org/10.1371/journal.pone.0211466.
[20] Z. Zhang, Microsoft kinect sensor and its effect, IEEE Multimed. 19 (2) (2012) 4–10, https://doi.org/10.1109/MMUL.2012.24.
[21] X. Zhou, X. Sun, W. Zhang, S. Liang, Y. Wei, Deep kinematic pose regression, in: Comput. Vis. – ECCV 2016 Work, 2016, pp. 186–201.
[22] V.-T. Hoang, K.-H. Jo, 3-D human pose estimation using cascade of multiple neural networks, IEEE Trans. Ind. Inform. 15 (4) (2019) 2064–2072.
[23] D. Guffanti, A. Brunete, M. Hernando, Development and validation of a ROS-based mobile robotic platform for human gait analysis applications, Robot. Auton. Syst. 145 (2021) 103869, https://doi.org/10.1016/j.robot.2021.103869, https://www.sciencedirect.com/science/article/pii/S0921889021001548, https://linkinghub.elsevier.com/retrieve/pii/S0921889021001548.

[24] D. Guffanti, A. Brunete, M. Hernando, J. Rueda, E. Navarro, ROBOGait: a mobile robotic platform for human gait analysis in clinical environments, https://doi.org/10.3390/s21206786, 2021.
[25] R.B. Davis, S. Õunpuu, D. Tyburski, J.R. Gage, A gait analysis data collection and reduction technique, Hum. Mov. Sci. 10 (5) (1991) 575–587, https://doi.org/10.1016/0167-9457(91)90046-Z, http://www.sciencedirect.com/science/article/pii/016794579190046Z.
[26] J.-C. Ceccato, M. de Sèze, C. Azevedo, J.-R. Cazalets, Comparison of trunk activity during gait initiation and walking in humans (trunk activity in walking), PLoS ONE 4 (12), https://doi.org/10.1371/journal.pone.0008193.
[27] M. Ma, R. Proffitt, M. Skubic, Y.-K. Jan, Validation of a Kinect V2 based rehabilitation game, PLoS ONE 13 (8), https://doi.org/10.1371/journal.pone.0202338.
[28] T. Wei, B. Lee, Y. Qiao, A. Kitsikidis, K. Dimitropoulos, N. Grammalidis, Experimental study of skeleton tracking abilities from Microsoft kinect non-frontal views, in: 3DTV-Conference, July 2015.
[29] G. Wu, P.R. Cavanagh, ISB recommendations for standardization in the reporting of kinematic data, J. Biomech. 28 (10) (1995) 1257–1261, https://doi.org/10.1016/0021-9290(95)00017-C.
[30] J.A. Zeni, J.G. Richards, J.S. Higginson, Two simple methods for determining gait events during treadmill and overground walking using kinematic data, Gait Posture 27 (4) (2007) 710–714.
[31] D.A. Winter, Biomechanics and Motor Control of Human Movement, 4th edition, John Wiley & Sons, Hoboken, New Jersey, 2009.
[32] O. Beauchet, C. Annweiler, Y. Lecordroch, G. Allali, V. Dubost, F.R. Herrmann, R.W. Kressig, Walking speed-related changes in stride time variability: effects of decreased speed, J. NeuroEng. Rehabil. 6 (1) (2009) 32.
[33] S.K. Challa, A. Kumar, V.B. Semwal, N. Dua, An optimized-LSTM and RGB-d sensor-based human gait trajectory generator for bipedal robot walking, IEEE Sens. J. 22 (24) (2022) 24352–24363, https://doi.org/10.1109/JSEN.2022.3222412.
[34] L. Konz, A. Hill, F. Banaei-Kashani, ST-DeepGait: a spatiotemporal deep learning model for human gait recognition, Sensors 22 (20) (2022), https://doi.org/10.3390/s22208075, https://www.mdpi.com/1424-8220/22/20/8075.
[35] H. Lim, B. Kim, S. Park, Prediction of lower limb kinetics and kinematics during walking by a single IMU on the lower back using machine learning, Sensors 20 (1) (2019), https://doi.org/10.3390/s20010130, https://www.mdpi.com/1424-8220/20/1/130.