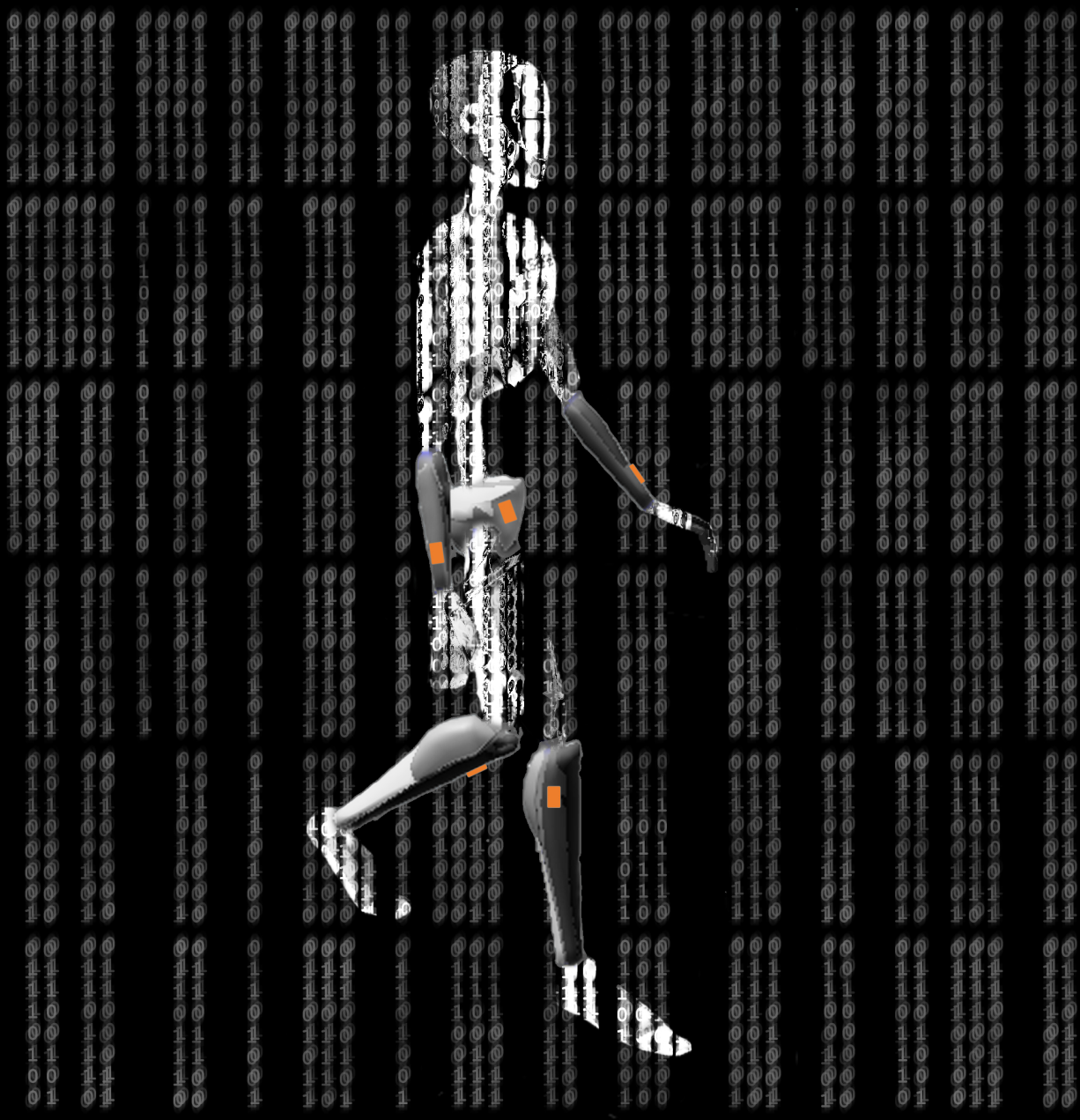


Towards Motion Capture with Minimal Sensing

Frank J. Wouda

Towards motion capture with minimal sensing

Frank Jasper Wouda

Department of Biomedical Signals and Systems, Faculty of Electrical Engineering, Mathematics, and Computer Science, Technical Medical Center, University of Twente

This research (project No.: 13917) is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

Paranymphs	Angelos Karatsidis and Kostas Konsolakis
Cover	Frank Jasper Wouda
Printing	Ipskamp
ISBN	978-90-365-4923-3
DOI	10.3990/1.9789036549233

Towards Motion Capture with Minimal Sensing

Dissertation

to obtain
the degree of doctor at the University Twente,
on the authority of the Rector Magnificus,
Prof.dr. T.T.M. Palstra,
on account of the decision of the graduation committee
to be publicly defended
on Friday 28 February 2020, at 12:45

by

Frank Jasper Wouda

born on 7 November 1990
in Hengelo

This dissertation has been approved by:

The supervisor:	Prof. dr. ir. P.H. Veltink
The co-supervisors:	Dr. ir. B.J.F. van Beijnum
	Dr. ir. M. Giuberti

Composition of the Graduation Committee:

Chairman & secretary:

Dean EEMCS

Supervisor:

Prof. dr. ir. P.H. Veltink	University of Twente

Co-supervisors:

Dr. ir. B.J.F. van Beijnum	University of Twente
Dr. ir. M. Giuberti	RADiCAL

Members - internal:

Prof. dr. J.N. Kok	University of Twente
Prof. dr. ir. J.H. Buurke	University of Twente

Members - external:

Prof. dr. ir. J. Harlaar	VUMC Amsterdam Delft University of Technology
Prof. dr. W. Zijlstra	Deutsche Sporthochschüle Köln
Prof. dr. ir. N.M. Maurits	University of Groningen

Summary

Human motion capture is important for a wide variety of applications, e.g., biomechanical analysis, virtual reality and character animation. Current human motion capture solutions require a large number of markers/sensors to be placed on the body. This results in long setup times and can be obtrusive for the subject. Publicly available motion capture datasets allow for using data-driven approaches to reduce the number of sensors/markers. While for specific applications an approach based on mechanical principles for sensor/marker reduction might be better suited. To that end, different approaches for sensor/marker reduction were developed and assessed, which were applied in specific applications.

The first chapter provides an overview of the most influential developments in the history of human motion analysis. Subsequently, different technologies used to measure both kinematics and kinetics in various measurement environments are presented. Next, methods for reducing the number of sensors or markers by using either data-driven or mechanical principles-driven approaches are described. Specific applications might benefit from an approach based on mechanical principles if appropriate assumptions can be applied. While a data-driven approach might be better suited if a more general solution is required. Followed by an overview of data-driven approaches that can be used to decrease the number of sensors. The goals of this thesis were to develop, optimize and evaluate methods for estimating full-body poses using a minimal sensor set and subsequently implement the proposed methods in relevant application scenarios.

The second chapter provides a comparison of the use of lazy and eager learning methods for estimation of full-body movements from a minimal sensor set. It was chosen to use nearest neighbor search as a lazy learning approach, as this could be fairly compared to previous work of one of the authors. The chosen eager learning approach was an artificial neural network, since it has been successfully applied to various problems that involve human motion. The results show that both learning approaches lead to similar estimation accuracy, namely an average joint position error of 7 cm and average joint angle error of 7 degrees. Additionally, both approaches show similar performance when input data is magnetically disturbed. This allows for applications that require real-time estimation of full-body poses, since eager learning approaches require significantly less processing time.

In the third chapter, improvements of the time coherency of output poses of the previously developed eager learning method are introduced by using a stacked input neural network. This approach was compared with a state-of-the-art deep neural network approach, which explicitly takes into account the time coherency of the in-/outputs. A shallow learning approach requires less computing power and data to train, which allows for applicability for a wide range of applications. Results

show that these deep and shallow learning approaches show comparable accuracy in estimation of full-body poses using only five inertial sensors, namely a joint position error of approximately 7 cm. However, an increase in the joint jerk errors was observed for the shallow approach compared to the deep learning approach, which was to be expected as time coherency was not explicitly enforced.

In the fourth chapter, the developed approach for reducing the number of sensors for full-body motion capture is applied to a running application. This was extended to estimate vertical kinetics of an athlete, which allows for a more detailed analysis of the athlete's performance. This approach was evaluated using data of a single subject and of multiple subjects, to analyze the generalizability of the proposed approach. As expected, a subject-specific trained network estimates kinematics and kinetics with a higher accuracy ($\rho > 0.99$) than a network trained on multiple subjects ($\rho > 0.9$). This performance can likely be improved by extending the training dataset to a more diverse and larger group of runners.

In the fifth chapter, an approach based on mechanical principles is applied for estimating the foot progression angle from a single foot-worn inertial sensor. By using a dynamic foot reference frame, there was no need for an external heading measurement, hence, a magnetometer is no longer required. Therefore, this approach is not affected by magnetic disturbances, and can be applied in any environment. Results show that the foot progression angle can be estimated with high accuracy compared to an optical reference (maximum mean error of 2.6°). This approach has the potential to be applied in a feedback application for knee osteoarthritis patients, to help them in reducing the knee adduction moment and consequently the experienced pain.

In the sixth chapter, different motion capture approaches are compared during running, namely: based on inertial measurement units (processed with Xsens MVN Analyze) and optical markers (processed using Plug-In Gait and OpenSim Gait2392). The evaluation was performed on a treadmill, with subjects running at 10, 12 and 14 km/h. The results show that mainly the sagittal plane has excellent correlation ($\rho > 0.96$) and RMSE (< 6 degrees). The transversal and frontal planes showed less correlation. This indicates that sagittal kinematics can be measured consistently using any of the different motion capture approaches, while ambiguities exist in the transversal and frontal planes even between same optical data processed with different accepted processing approaches.

Finally, in the last chapter, the insights and limitations of the research presented in this thesis are presented. First steps towards reducing the number of sensors for full-body motion capture have been taken in this work, however, improvements are required for these techniques to be applied in different applications. This provides opportunities for applications that require truly ambulatory motion capture and cannot apply a large set of sensors, e.g., for daily-life monitoring of patients or athletes. The accuracy of a data-driven approach is largely impacted by the dataset quality and size. Therefore, improvements can be achieved in this work using a more diverse and larger dataset. Further performance improvements can be realized by fine-tuning the proposed approaches to specific applications, such as was done for running.

In the appendix, a magnetic measurement system for estimation of end-effector position for a virtual reality application to train reaching tasks in stroke survivors is evaluated. An active magnet was used such that a 3D position/orientation can

be estimated without requiring additional information and/or assumptions. Results show that the end-effector position is estimated with a difference of approximately 3 cm compared to an optical reference. This demonstrates that the evaluated magnetic measurement system has potential to be applied in such a rehabilitation application. However, the developed mock-up for the virtual reality training application requires further improvement and subsequent experimental evaluation in stroke subjects.

Samenvatting

Vastleggen van menselijk bewegen is belangrijk voor een grote verscheidenheid aan toepassingen, bijvoorbeeld biomechanische analyses, virtuele realiteit en karakter animatie. Huidige oplossingen voor het vastleggen van menselijk bewegen vereisen dat een groot aantal reflectoren/sensoren op het lichaam worden geplaatst. Dit resulteert in lange voorbereidingstijd en kan de gebruiker in de weg zitten. Publiek beschikbare datasets met bewegingsdata maken het mogelijk om een data-gedreven aanpak toe te passen om het aantal reflectoren/sensoren te minimaliseren, maar voor specifieke toepassingen kan een aanpak die gebruik maakt van mechanische principes beter geschikt zijn. Daarom zijn er verschillende aanpakken voor reflector/sensor reductie geïmplementeerd en geëvalueerd, welke in specifieke applicaties zijn toegepast.

Het eerste hoofdstuk geeft een overzicht van de meest invloedrijke ontwikkelingen in de geschiedenis van de analyse van het menselijk bewegen. Vervolgens, worden er verschillende technologiën voor het meten van kinetica en kinematica in verscheidene omgevingen gepresenteerd. Daarna, worden methodes voor het reduceren van het aantal sensoren of reflectoren door gebruik te maken van data-gedreven of mechanische principes-gedreven aanpakken beschreven. Specifieke toepassingen hebben mogelijk baat bij een aanpak gebaseerd op mechanische principes wanneer de juiste aannames toegepast kunnen worden. Terwijl een data-gedreven aanpak mogelijk beter toegepast kan worden, wanneer een meer generieke oplossing nodig is. Vervolgens wordt er een overzicht van data-gedreven aanpakken gegeven die toegepast kunnen worden om het aantal sensoren te reduceren. De doelen van dit proefschrift waren: het ontwikkelen, optimaliseren en evalueren van methodes voor het gebruik van een minimale sensor set voor het schatten van bewegingen van het hele lichaam en vervolgens deze methodes toepassen in relevante applicatiescenario's.

In het tweede hoofdstuk wordt een vergelijking gemaakt tussen het gebruik van lazy en eager learning methodes voor het schatten van bewegingen van het hele lichaam met een minimale sensor set. Er is gekozen om nearest neighbor search als lazy learning methode te gebruiken, omdat deze eerlijk met eerder werk van één van de auteurs vergeleken kan worden. De gekozen eager learning methode was een neuraal netwerk, omdat deze succesvol is toegepast op verschillende problemen met bewegingsdata. De resultaten laten zien dat beide methodes tot een vergelijkbare schattingsnauwkeurigheid leiden, namelijk een gemiddelde gewrichtspositiefout van 7 cm en een gemiddelde gewrichtshoekfout van 7 graden. Verder laten beide methodes vergelijkbare prestaties zien wanneer de input data magnetisch verstoord is. Dit maakt applicaties mogelijk waarbij onvertraagde schattingen van lichaamsposes nodig zijn, omdat eager learning significant minder verwerkingstijd nodig heeft.

In het derde hoofdstuk worden verbeteringen van de tijdscoherentie van de ge-

schatte poses van de eerder gepresenteerde eager learning methode geïntroduceerd door gebruik te maken van een gestapeld input neuraal netwerk. Deze aanpak is vergeleken met een state-of-the-art deep neuraal netwerk aanpak, welke expliciet de tijdscoherentie van de in/outputs in acht neemt. Een shallow learning aanpak heeft minder computerkracht en trainingsdata nodig, dit maakt toepassing voor een groot bereik van applicaties mogelijk. De resultaten laten zien dat deep en shallow learning methodes vergelijkbare nauwkeurigheid hebben in het schatten van lichaamsposes met maar vijf inertiële sensoren, namelijk een gewrichtspositiefout van ongeveer 7 cm. Hoewel er wel een stijging in de gewrichtsschokfout te zien is voor de shallow aanpak in vergelijking met de deep learning aanpak, wat te verwachten was omdat hierbij tijdscoherentie niet expliciet gemodelleerd was.

In het vierde hoofdstuk is de ontwikkelde aanpak voor het reduceren van het aantal sensoren voor het vastleggen van menselijk bewegen toegepast in een hardloop applicatie. Dit was uitgebreid met een schatting van de verticale kinetica van de atleet, dit zorgt ervoor dat een uitgebreidere analyse van prestaties van de atleet mogelijk is. Deze aanpak is geëvalueerd met data van één proefpersoon en meerdere proefpersonen, om zo de generaliseerbaarheid van de aanpak te kunnen analyseren. Zoals verwacht is een netwerk dat is getraind in het schatten van de kinetica en kinematica van één proefpersoon nauwkeuriger ($\rho > 0.99$) dan een netwerk getraind met verschillende proefpersonen ($\rho > 0.9$). Deze prestaties kunnen waarschijnlijk nog verbeterd worden door de training dataset uit te breiden naar een meer diverse en grotere groep hardlopers.

In het vijfde hoofdstuk is een aanpak gebaseerd op mechanische principes toegepast op het schatten van de voet verloophoek met één voet-gedragen inertiële sensor. Door gebruik te maken van een dynamisch voet referentie assenstelsel was er geen externe meting nodig voor de horizontale richting, oftewel een magnetometer is niet langer noodzakelijk. Omdat deze aanpak niet beïnvloed wordt door magnetische verstoringen kan dit in iedere omgeving worden toegepast. De resultaten laten zien dat de voet verloophoek geschat kan worden met een hoge nauwkeurigheid in vergelijking met een optische referentie (maximale gemiddelde fout van 2.6°). Deze aanpak heeft de potentie om toegepast te worden in een feedback applicatie voor knie osteoarthritis patiënten, zodat zij ondersteund kunnen worden in het verlagen van hun knie adductie moment en derhalve de ervaarde pijn.

In het zesde hoofdstuk zijn verschillende aanpakken voor het vastleggen van bewegingen vergeleken tijdens het hardlopen, namelijk gebaseerd op inertiële sensoren (verwerkt met Xsens MVN Analyze) en optische reflectoren (verwerkt met Plug-In Gait en OpenSim Gait2392). De evaluatie was uitgevoerd op een loopband, met proefpersonen hardlopend met een snelheid van 10, 12 en 14 km/h. De resultaten laten vooral in het sagittale vlak excellente correlatie ($\rho > 0.96$) en RMSE ($< 6°$) zien. De transversale en frontale vlakken laten een lagere correlatie zien. Dit geeft aan dat sagittale kinematica consistent met elke van de verschillende aanpakken gemeten kan worden, terwijl dit niet eenduidig is voor het transversale en frontale vlak, zelfs wanneer men kijkt naar optische data verwerkt met verschillende geaccepteerde methodes.

Tot slot, in het laatste hoofdstuk worden inzichten en beperkingen van het onderzoek in dit proefschrift gepresenteerd. De eerste stappen in de richting van het reduceren van het aantal sensoren voor het vastleggen van bewegingen zijn gezet in dit proefschrift, hoewel er nog wel verbeteringen nodig zijn om deze technieken

toe te kunnen passen in verschillende applicaties. Dit biedt mogelijkheden voor toepassingen die vereisen dat de bewegingen vastgelegd worden in een ambulante setting en wanneer het niet mogelijk is om een grote set sensoren toe te passen, bijvoorbeeld voor het monitoren van patiënten of atleten tijdens het dagelijks leven. De nauwkeurigheid van een data-gedreven aanpak wordt grotendeels beïnvloed door de kwaliteit en grootte van de gebruikte dataset. Daarom kunnen de methodes in dit werk nog verbeterd worden door een meer diverse en grotere dataset te gebruiken. Daarnaast kan er nog verbetering gerealiseerd worden door de beschreven aanpakken toe te spitsen op specifieke toepassing, zoals bijvoorbeeld voor de hardloop applicatie is gedaan.

In de bijlage is een magnetisch meetsysteem getest waarmee de eindeffector positie geschat kan worden voor een virtuele realiteit applicatie voor het trainen van reikbewegingen na een beroerte. Er is een actieve magneet gebruikt zodat de 3D positie en oriëntatie geschat kan worden zonder dat hiervoor additionele informatie en/of aannames nodig zijn. De resultaten laten zien dat de eindeffector positie geschat kan worden met een maximaal verschil van ongeveer 3 cm in vergelijking tot een optische referentie. Dit laat zien dat dit geteste magnetische meetsysteem potentie heeft om toegepast te worden in zo'n revalidatie applicatie. Maar het ontworpen testmodel voor de virtuele realiteit trainingsapplicatie moet eerst verder verbeterd worden en vervolgens geëvalueerd worden met patiënten die een beroerte hebben gehad.

Table of Contents

List of Abbreviations

A/A	Ab-/Adduction
ADL	Activity of Daily Living
ANN	Artificial Neural Network
BW	Body Weight
F/E	Flexion/Extension
FP	Forceplate
FPA	Foot Progression Angle
I/E	Internal/External Rotation
IMMU	Inertial Magnetic Measurement Unit
IMU	Inertial Measurement Unit
KAM	Knee Adduction Moment
KOA	Knee Osteoarthritis
LSTM	Long Short-Term Memory
NNS	Nearest Neighbor Search
OS	OpenSim
PCA	Principal Component Analysis
PiG	Plug-in-Gait
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SIL	Stacked Input Length
SINN	Stacked Input Neural Network
STA	Soft-Tissue Artifact
vGRF	Vertical Ground Reaction Force
ZAVU	Zero Angular Velocity Update
ZUPT	Zero Velocity Update

Chapter

ONE

Introduction

Human motion analysis is important for various research areas, e.g. sports, rehabilitation, design of orthotic and prosthetic devices, etc. Currently such analysis is performed by using either a lab-bound or portable measurement system, each with their own benefits and limitations. However, for full-body motion capturing these systems often require sensors/markers on each rigid-body segment. This results in long set up times and can be obtrusive to the user. Availability of motion capture data in public databases, provides opportunities for applying machine learning to decrease the number of sensors/markers. However, in case of specific applications the use of mechanical principles might result in better estimation of quantities of interest. In this thesis we applied both machine learning and mechanical principles for estimating full-body movements. Additionally, both methods have been applied to specific applications. This chapter provides an overview of the background of different motion analysis, methods for reduction of sensor/markers and machine learning.

1.1 Motion Analysis

Interest in the movement of humans goes back to the earliest times. However, Aristotle (384-322 BCE) is the author of the first known written reference in this research field [17], where he suggested the following experiment for measuring human gait:

> "If a man were to walk on the ground alongside a wall with a reed dipped in ink attached to his head the line traced by the reed would not be straight but zig-zag, because it goes lower when he bends and higher when he stands upright and raises himself."[12]

Unfortunately, it took until the development of modern mathematics in the 16th century to continue research into human movement and perform the appropriate measurements. The first gait experiment was performed by Giovanni Alfonso Borelli (1608–1679), in which he walked towards two poles placed at unknown distance, while trying to keep only one pole in sight [26]. However, he found that this was impossible due to medio-lateral movement of the head. Measuring human movement requires proper equipment, or else one might come to incorrect conclusions. This happened in the work of Weber and Weber (1836), who relied on measuring tape, a stopwatch and telescope, in combination with their observations to draw 14 instances (in time) of a gait cycle [203]. Their results were proven to be not completely accurate by the first three-dimensional gait analysis performed by Braune and Fisher (1891, [29]). In this analysis subjects were strapped with Geissler tubes that flashed at discrete intervals, which was recorded with a continuous exposure, the setup is shown in Figure 1.1 and results in Figure 1.2.

Besides kinematics, it is important to measure kinetics to understand the full scope of forces exerted on the human body during motion. The first researcher to measure the ground reaction forces was Gaston Carlet (1849–1892) using a shoe that had pressure transducers built-in (shown in Figure 1.3, with corresponding results in Figure 1.4) [32]. This technology was adapted to be applied in the first pneumatic (vertical) force plates [123].

However, to understand the full picture of forces acting on the body all three orthogonal components of the ground reaction forces are required. To that end, Jules Amar (1879-1935) developed such a device during the first World War (shown in Figure 1.5) capable of measuring three-dimensional ground reaction forces [5].

The introduction of a force plate in combination with motion capture provided all the necessary elements for performing adequate full-body three-dimensional gait analysis. However, processing of those data required significant human effort, as all calculations had to be performed by hand. It took until the computer era for gait analysis to be available to a wider range of labs. Where the first important innovation was capturing positions of retro-reflective markers (accuracies that can be reached are in the order of millimeters [64]) using a calibrated setup of infra-red cameras, e.g., as shown in Figure 1.6. Furthermore, this optical motion capture system can be complemented with force plates, such as shown in Figure 1.7. When retro-reflective markers are placed on known body locations, three-dimensional kinematics can be inferred. To obtain segment orientations three markers have to be placed on each rigid body segment with high accuracy, as a small misplacement

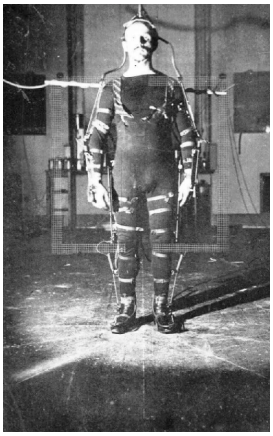


Figure 1.1: The first known three-dimensional motion capture system consisting (designed by Braune [29]) of Geissler tubes attached to a suit with thick layers of rubber (to protect against the electrical current). It took between 6 and 8 hours to dress the subject, and measurements could only be performed in complete darkness (i.e., at night).

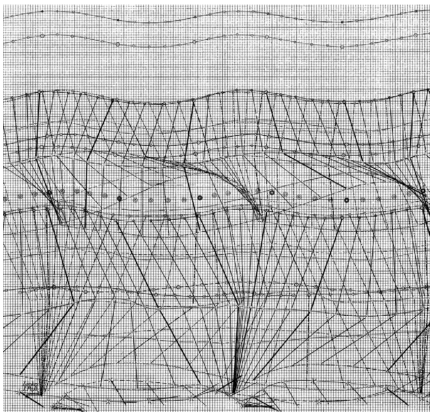


Figure 1.2: After careful processing of the acquired images from recordings of a subject wearing the motion capture system (as shown in Figure 1.2) the joint centres were plotted in this mesh. Additionally, centre of mass calculations were performed and included in this representation [29].

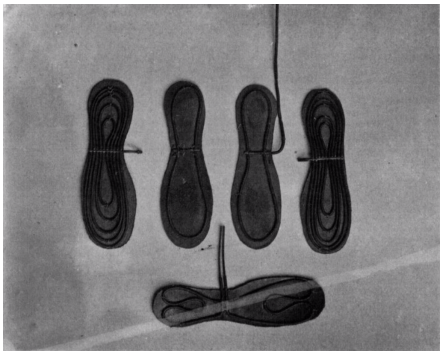


Figure 1.3: Pressure transducers built in soles, such that ground reaction forces could be measured [32, 123].

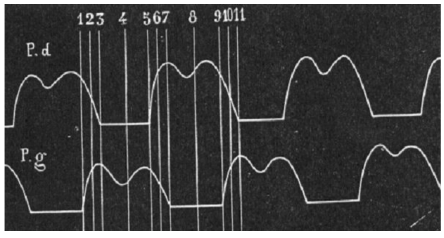


Figure 1.4: The first recorded characteristic double bump of the vertical ground reaction force (by Gaston Carlet) [32].

might result in a significant segment orientation error. Furthermore, it has been shown that this placement is subject to inter-session variability, even when placed by the same researcher on the same subject [124].

Optical motion capture requires significant setup and processing times due to the large number of markers (37 if the Vicon Plug-In Gait protocol is used [49, 201]) that need to be placed and consequently processed. This processing consists of labeling the observed retro-reflective markers (which can be automated with the appropriate model, but can result in label errors) and the resulting marker trajectories can require gap-filling (due to occlusions). Furthermore, a dedicated area with calibrated

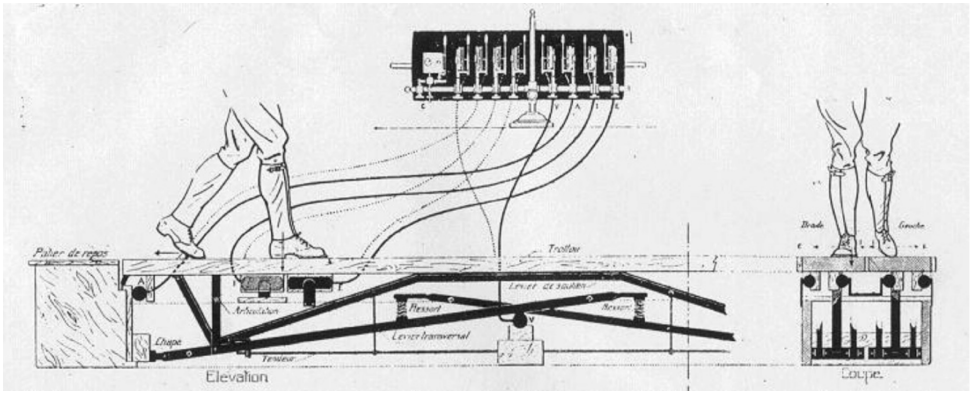


Figure 1.5: The first force plates capable of measuring three-dimensional forces [5].

Figure 1.6: An optical motion capture system setup of Vicon [201].

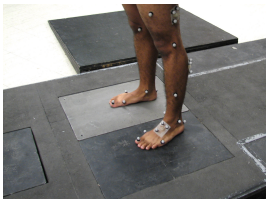


Figure 1.7: Optical motion capture in combination with force plates [158].

cameras (i.e., positions of cameras with respect to each other should be known) is required, which limits the applicability to lab measurements or requires significant effort to set up the system in a different environment (e.g., outdoors or in a home). Even though optical motion capture is regarded the golden standard for measuring human kinematics, it has been shown that due to its' limitations (e.g., lab-based measurements, line of sight issues and post-processing requirements) other approaches might be a better choice depending on the measurement conditions [42, 208].

Inertial motion capture on the other hand, uses on-body sensors, namely inertial measurement units (IMUs, accelerometer, gyroscope and magnetometer) applied to each (rigid) body segment (17 in total) to determine orientation of the various body segments [170], as shown in Figures 1.8 and 1.9. This has two major advantages compared to optical systems, namely that it is not bound to a dedicated lab space and does not suffer from occlusions. This significantly reduces the setup times, and with no occlusions gap-filling is not required. An IMU (as shown in Figure 1.9) typically consists of a three-dimensional accelerometer and gyroscope, which is often complemented with a magnetometer.

An accelerometer can be used to measure the acceleration that is acting on the sensor in all directions, which includes the gravitational acceleration. However, in essence it measures the forces acting on the mass attached in a box via springs in all three perpendicular directions, by evaluating Hooke's law ($F = kx$) using the measured displacement of the mass inside the sensor box (x) and known spring stiffness

Figure 1.8: Inertial motion capture system of Xsens [214].

Figure 1.9: Wireless IMUs of Xsens [214].

(k). Newton's law ($F = ma$) is then applied to estimate acceleration (a) using the calculated force (F) and known mass (m). To remove the gravitational acceleration from this measured acceleration, orientation of the sensor must be known in an Earth reference frame.

This information is typically combined with a rate gyroscope and magnetometer, using a sensor fusion technique such as an Extended Kalman Filter [164], to obtain the orientation of the sensor w.r.t. the Earth reference frame. A rate gyroscope was initially based on a rotating structure, however, a vibrating structure gyroscope is simpler and cheaper and therefore more suited to be used in MEMS technology. These type of gyroscopes are based on a vibrating mass that exerts a force on its supporting structure according to the Coriolis effect. Then the rate of angular rotation can be calculated using the measured forces in the three perpendicular directions. A magnetometer is a device that measures the magnetic field (direction and strength) in a certain location. This can be used to calculate the heading direction of the device using knowledge of the Earth magnetic field (like a compass). However, this requires minimal magnetic disturbances (e.g., ferro-magnetic material or power sources in the vicinity of the device). With good tuning of an Extended Kalman Filter, these three information sources can be combined to estimate an accurate and drift-free three-dimensional orientation [168]. This can then be used to calculate the free acceleration of the sensor (without the gravitational acceleration component), and transform the measured accelerations into the global frame.

With a proper biomechanical model it becomes possible to acquire body segment orientations and positions. This requires knowledge about the placement of IMUs on the different body segments, which can be obtained by a calibration procedure in which the subject is in a known pose [161]. In this manner, proper sensor-to-segment alignment can be acquired assuming that IMUs are rigidly attached to the human body and an accurate calibration pose was performed. If segment lengths are measured accurately, joint positions can be estimated using forward kinematics [13, 160].

A significant limitation of inertial motion capture compared to optical motion capture is drift in position of the subject, w.r.t. the Earth reference frame, due to double numerical integration of the accelerations. This position is typically estimated by inertial dead-reckoning using the known initial position and velocity, which results

in drift due to accumulation of measurement errors and sensor noise. Effects of this drift can be minimized by applying additional information from interactions with the environment, e.g., zero velocity or height updates during stance in walking and running [183].

Until recently, inertial and magnetic motion capture suffered from magnetic disturbances [50]. This was caused by the assumption that the magnetometer measures a homogeneous magnetic field, which is not true in most indoor environments due to the presence of ferromagnetic materials in a lot of buildings. However, a recent improvement to the algorithms allows for accurate estimation of kinematics even in the presence of external factors, e.g., magnetic disturbances [170]. Also in specific cases, three-dimensional movements can be analyzed using inertial sensors without the need for additional magnetic sensing.

1.2 Minimal Sensing

All previously mentioned motion capture technologies, require sensors or markers on each (rigid) body segment. This results in long setup times and can be obtrusive to the user. Human motion is regarded highly redundant, therefore, reduction of this number of sensors/markers is deemed possible [165, 166, 195]. There are two significantly different approaches to achieve such a sensor/marker reduction, namely by taking advantage of knowledge about human kinematics in relation to the underlying physical principles (e.g., [82]) or by inferring such information from a large dataset of human motion (e.g., [14]).

1.2.1 Mechanical Principles Modeling

Full-body motion capture requires applying assumptions about human body motion, e.g., joint constraints, limitations on speed of movement or joint type [170]. However, additional assumptions can be made to reduce the number of sensors for (full-body) motion capture, which may result in a decrease of accuracy, but might be effective for specific applications.

A special case of a reduction of the number of available markers is when these are occluded during (part of) a measurement. Such cases can be solved using interpolation (over typically short periods of time) with information from the remaining markers. Such interpolations can effectively be achieved using linear motion modeling with a Kalman Filter [11] or by exploiting knowledge of rigid body segments of the human body [80, 190].

Specific applications might benefit from additional constraints that allow for minimizing the number of markers/sensors for estimation of full-body kinematics, e.g., 8-marker model for biofeedback of gait training [25] or 8 IMUs for tracking changes of running mechanics during a marathon [149]. The number of markers and/or setup can be optimized for the specific application, however, it requires different assumptions and models to estimate required outcomes with sufficient accuracy. A different approach is to use optimization in combination with a biomechanical model for full-body motion capture with a set of 6 IMUs [202].

1.2.2 Data-driven Modeling

The increasing availability of human body motion datasets (e.g., CMU [33], AMASS [122] and HDM [129]) provides opportunities to create data-driven models to map a low-dimensional marker and/or sensor setup to a full-body pose. Inferring such relations based on data can result in acceptable performance, but might not provide additional insights into the underlying relations as such machine learning approaches are typically regarded to be "black-box" methods (this might change with increased interest in explainable artificial intelligence [1]). However, it has been shown to be an effective class of tools to reduce the number of markers/sensors in full-body motion capture, e.g., using minimal retro-reflective markers [36] or IMUs [88, 192, 210].

1.3 Machine Learning

It is becoming increasingly important to effectively use the growing amount of data (in various domains), as we might gain and infer new insights and knowledge that is implicitly present or encoded in these data. To that end, machine learning has been applied with varying success for classification and/or regression based on such data. These learning methods can be unsupervised (when no labeled data is available) or supervised (when labeled data is available). However, in the following focus is on supervised learning methods.

Machine learning algorithms can be divided in two main categories, namely lazy and eager learning [3]. If the algorithm combines information from a data set directly to process new observations and does not learn from it, we can classify it as a lazy learning algorithm. This also means that no model is created to relate in- and outputs. Examples of such algorithms are nearest neighbour search and naive Bayes. In eager learning a model is constructed based on the available data that learns relations between the respective in-/outputs, e.g., artificial neural networks, decision trees, etc. Both paradigms have their own advantages, namely lazy learning does not require training time, which comes at the expense of processing time at runtime and eager learning does not require the full dataset to be stored, because this knowledge is contained in the trained model, which makes it faster at runtime.

There are numerous algorithms that can be seen as machine learning [22]. However, to stay within the scope of this thesis we will discuss artificial neural networks (ANN) only. ANNs are inspired by the biological neural networks present in animals. In general they learn to perform tasks (regression or classification) considering training examples using no task-specific rules, i.e., the brain can be used for many different tasks using the same structure. An ANN is typically structured with one input layer that is fully-connected to one or two hidden layers which are fully-connected to an output layer, as shown in Figure 1.10. As deeper ANN (more hidden layers) are difficult to train, because of the vanishing gradient problem [83]. Each layer consists of neurons (the computing units of the network), which typically outputs a non-linear sum of its inputs. The number of neurons in the hidden layers can be determined experimentally. Weights of the different neurons in the network are trained iteratively by evaluating a cost function on the available training examples of inputs and corresponding outputs. If this cost function is minimized or the pre-defined number of iterations is achieved, the training phase is finished and the ANN can be used to

calculate estimates of the outputs based on only inputs.

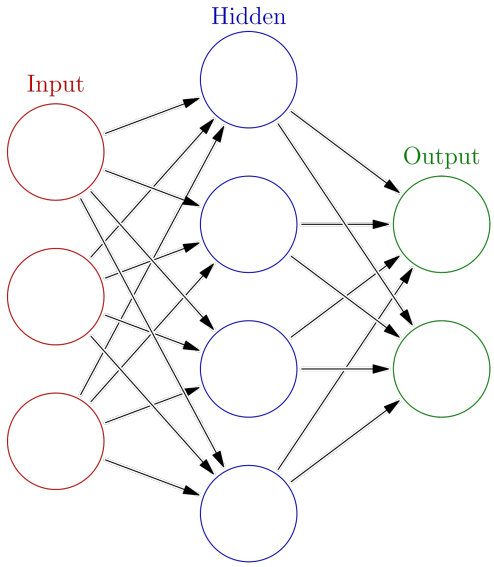


Figure 1.10: Artificial neural network structure [204].

An important step in applying machine learning algorithms is feature engineering, i.e., extract information from a dataset in the form that is most relevant for the regression/classification problem at hand. The introduction of deep neural networks (more hidden layers) made feature engineering less important, since these additional layers can effectively transform the input data into relevant features [117]. This has been successfully applied to visual object and speech recognition [53]. However, deep learning requires significant computing power compared to shallow learning (neural networks with only one or two hidden layers). Therefore, choosing either deep or shallow learning depends on the available human and/or computing resources.

1.4 Thesis Objectives

This research has two objectives, which have been formulated as follows:

- Develop, optimize and evaluate a method for estimating full-body poses using a minimal sensor set.
- Implementation of the proposed methods in relevant application scenarios.

1.5 Thesis Outline

Based on these thesis objectives the following three parts, with their corresponding chapters, are defined as follows:

- **Methods for Pose Estimation with Minimal Sensing** - This part describes methods for reducing the number of sensors in general full-body pose estimation.

  2. **Estimation of Full-Body Poses Using Only Five Inertial Sensors: An Eager or Lazy Learning Approach?** - This chapter provides a comparison of two learning paradigms (e.g., eager and lazy learning) for estimating full-body poses with only five IMUs.
  3. **Time Coherent Full-Body Poses Estimated Using Only Five Inertial Sensors: Deep versus Shallow Learning** - Time coherent output was not achieved by using traditional shallow learning (Chapter 2), to that end, we developed a stacked input neural network, which was compared to the state-of-the-art deep learning approach for estimating full-body poses from a minimal sensor set.

- **Applications for Pose Estimation with Minimal Sensing** - This part describes applications for estimating human kinematics/kinetics with a small sensor set using methods from Part I or mechanical principles.

  4. **Estimation of Vertical Ground Reaction Forces And Sagittal Knee Kinematics During Running Using Three Inertial Sensors** - Describes how an ANN can be used effectively to estimate both the kinematics and kinetics of running, which is relevant for monitoring performance progression of the athlete and to possibly assess injury risks.
  5. **Foot Progression Angle Estimation Using A Single Foot-Worn Inertial Sensor** - Knee osteoarthritis patients can have mobility issues due to the experienced knee pain. Providing feedback on the foot progression angle can help this population to reduce their knee pain. This work provides a method to estimate the foot progression angle using a single foot-worn inertial sensor.

  A. **Magnetic Sensing For Estimation of End-Effector Position in a Virtual Reality Application** - A study of the potential for using minimal magnetic sensing for estimation of end-effector position in a virtual reality application. Which includes a preliminary design for such an application.

- **Technologies for Motion Capture** - This part provides a performance comparison between different motion capture technologies used in parts I and II.

  6. **On the Validity of Different Motion Capture Technologies For the Analysis of Running** - Optical and inertial motion capture technologies each have their benefits and limitations, however, this work provides insight in differences of obtained kinematics.

Chapter 7 concludes and discusses the obtained results of this work and identifies opportunities for future research directions.

Part I

Methods for Pose Estimation with Minimal Sensing

Chapter

TWO

Estimation of Full-Body Poses Using Only Five Inertial Sensors: An Eager or Lazy Learning Approach?

Published as:
F. J. Wouda, M. Giuberti, G. Bellusci and P. H. Veltink **Estimation of Full-Body Poses Using Only Five Inertial Sensors: An Eager or Lazy Learning Approach?** Sensors 2016, 16, 2138 https://doi.org/10.3390/s16122138

Abstract

Human movement analysis has become easier with the wide availability of motion capture systems. Inertial sensing has made it possible to capture human motion without external infrastructure, therefore allowing measurements in any environment. As high-quality motion capture data is available in large quantities, this creates possibilities to further simplify hardware setups, by use of data-driven methods to decrease the number of body-worn sensors. In this work, we contribute to this field by analyzing the capabilities of using either artificial neural networks (eager learning) or nearest neighbour search (lazy learning) for such a problem. Sparse orientation features, resulting from sensor fusion of only five inertial measurement units with magnetometers, are mapped to full-body poses. Both eager and lazy learning algorithms are shown to be capable of constructing this mapping. The full-body output poses are visually plausible with an average joint position error of approximately 7 cm, and average joint angle error of 7 °. Additionally, the effects of magnetic disturbances typical in orientation tracking on the estimation of full-body poses was also investigated, where nearest neighbour search showed better performance for such disturbances.

Keywords: Inertial Motion Capture; Orientation Tracking; Machine Learning; Neural Networks; Nearest Neighbor Search; Human Movement; Reduced Sensor Set.

2.1 Introduction

Human motion has been a research topic of interest in many fields for a long time. The increasing availability of high-quality motion capture systems [145, 146, 161, 201] contributed to this topic, by allowing easier and more accurate three-dimensional human motion capturing [70]. The introduction of inertial motion capture systems, which do not rely on any external infrastructure, made full-body movement analysis feasible in an ambulatory setting [161]. These systems require sensors to be attached to each main body segment (e.g., 17 sensors in Xsens MVN [214]). By reducing the number of required body-worn sensors, such systems would be less obtrusive and the usability would improve, which could potentially lead to applications that require use in daily life. Another probable benefit would be the reduction in costs.

Many studies have proved that human movements contain redundant information and can be concisely described using fewer dimensions than the degrees of freedom of the human body [165, 166, 195]. This opens the way for human motion capture using a reduced set of sensors. The first approaches in this field used procedural models (based on empirical and biomechanical models, they offer less control but are not perceived as realistic [130]). Badler et al. [14] used four magnetic sensors (placed on the head, pelvis and both hands) and inverse kinematics to estimate upper body joint positions. Behavioral models were used for the estimation of the lower body joint positions, which resulted in the best estimation of gait poses. Another example of a heuristic-based system used eight magnetic sensors to estimate full-body movements by applying an analytical solution for the authors' defined kinematic chains problem [174].

The increasing availability of motion capture data led to the use of data-driven approaches to deal with less information than provided by current full-body motion capture systems. One of the first data-driven approaches was presented by Chai and Hodgins [36], who used six reflective markers (captured with two video cameras) providing the position of anatomical landmarks to estimate full-body movements. A form of Nearest Neighbor Search (NNS) was used to map the lower dimensional input signals to full-body poses. The search space was limited with Principal Component Analysis to create a local linear model. The use of position-based features particularly fits with methods such as NNS, but calls for an external infrastructure (such as cameras) which limits the applicability to small (mostly indoor) areas.

In later years, Slyper and Hodgins [185] investigated a system composed of five accelerometers attached to the upper-body. Their results were promising and encouraged others to further investigate the use of such sensors for this problem. Tautges et al. [192] built upon that work by also using a few accelerometers, but with the addition of scaling in the temporal domain to also enable estimation of movements performed at different speeds. Riaz et al. [152] added ground contact information to the estimation framework of Tautges et al. [192] to estimate full-body poses using three accelerometers (placed on the wrists and lower back). These works showed the effort of moving towards an infrastructure-less setting, but the use of raw accelerometer data limited the potential performance of such methods, as these provide information of movement but not of a single pose, unlike position features.

The above-mentioned methods all adopted some form of NNS in their estimation framework. NNS belongs to the family of lazy learning algorithms [3], since

no model is created during a training phase. Although these algorithms have been shown to be a good choice for the estimation of full-body movements using a reduced sensor set [36, 96, 185, 192], the fact that the (typically large) training database needs to be stored makes them less appealing for real-time portable (or even embedded) applications. Even though smart search techniques (such as k-dimensional trees [36, 108, 192]) have been proposed to decrease search time, training databases cannot be indefinitely large to search and estimate output in real-time. Eager learning algorithms, opposed to lazy learning ones, do not have this requirement and a model is trained to concisely describe relationships between inputs and outputs in a training database. Liu et al. [120] applied a combination of both an eager and lazy learning approach to estimate full-body movements using six inertial and ultrasound sensors. Their approach combined previously estimated poses and their neighbors to construct a local linear model for prediction of the current pose. Starting from a similar setup, Kim et al. [101] used a kernel canonical-correlation-analysis-based regression. Although both works reported promising results with the use of eager learning approaches, their methods required several models to be trained at different time steps, thus increasing computational complexity. Moreover, both methods required position-based features and showed dependency on an external infrastructure.

A few examples of training global models for human movements can also be found in literature. For instance, support vector regression has been applied to the 2-Dimensional (2D) upper body pose estimation from images [40] and Artificial Neural Networks (ANNs) have been applied for the estimation of 3-Dimensional (3D) movements from 2D body poses [196]. None of these works, however, focus on estimating full-body movements from a reduced set of body-worn sensors.

In the efforts of developing a self-contained ambulatory system, preliminary work (chapter 5 of the PhD thesis) by one of the authors [74] has shown the feasibility of estimating full-body poses using an NNS-based approach with orientation features. However, this method also implemented simulated features, which would require additional sensor technologies. The choice for orientation features was mainly driven by the fact that current orientation tracking solutions [161], based on sensor fusion of inertial and magnetic data, have been proved to be highly effective at providing users with easy access to accurate and very informative quantities such as sensor orientations. Furthermore, orientations directly provide information of the current body pose, unlike only accelerations. In this work, we aim at further investigating the effectiveness of using orientation features in combination with a lazy learning (NNS) and an eager learning (ANN) algorithm. Since the main focus of the work is to investigate two learning paradigms (i.e., lazy and eager), the algorithm choice has been mainly driven by consistency with previous literature. Note that the choice between a lazy and eager learning approach shows a functional trade-off. Lazy learning, by not building any specific model, is more effective in preserving idiosyncrasies of training data. However, for the same reason, the computational complexity at run-time grows with the size of data. On the other hand, eager learning requires much longer training time, but the conciseness of the trained model makes it more appealing to real-time applicability. A data collection campaign has been carried out to create a large training database composed of movements from different subjects performing various activities and used to test the two algorithms. The sensor orientations provided by five Xsens motion trackers [214] (each containing an Inertial

Measurement Unit (IMU) and magnetometer) have then been used as inputs to the learning algorithms to estimate full-body poses. Our method enables the realistic estimation of full-body poses using a reduced set of IMUs.

The remainder of this paper is organized as follows. In Section 2.2, we describe the data collection and processing. In Section 2.3, the detailed performance of both algorithms for different configurations and activities, and the effect of magnetic disturbances are described and discussed. Finally, conclusions of this work are presented in Section 2.4, and possible future work is described in Section 2.5.

2.2 Methods

2.2.1 Subjects

Six healthy subjects (three males and three females; age: 29.0 ± 11.9 years; height: 177.2 ± 7.2 cm; weight: 83.1 ± 11.1 kg; Body Mass Index (BMI): 26.5 ± 3.7 kg/m^2; all dominant right-handed) volunteered to participate in the current study, for which ethical approval was obtained. Before participating in this research, each subject signed an informed consent form.

2.2.2 Experimental Protocol

Subjects were verbally instructed to perform different types of movements (ranging from gait, Activities of Daily Living (ADLs), to sports), as described in Table 2.1. The exact execution of these movements (namely, style and speed) was left to individual interpretation. Each trial was performed three times to guarantee large intra- and inter-subject variations in the movement database (which can be seen from some representative recordings shown in accompanying videos). Overall, approximately 25 min of motion capture data was recorded for each subject. The measurements were performed in the gait laboratory of Roessingh Research and Development (Enschede, the Netherlands). The working area used for data collection was chosen trying to minimize magnetic disturbances (e.g., large metal constructions were avoided and ferro-magnetic objects, when possible, were removed from the working area), such that good quality measurements were possible. MVN Studio offers some dedicated tools to measure magnetic field experienced in the environment and to test its homogeneity.

2.2.3 Instrumentation

The reference full-body motion capture system used in this study is Xsens MVN (Xsens Technologies B.V., Enschede, the Netherlands), in this manner, both learning approaches could be compared to the performance of an ambulatory motion capture system. Xsens MVN consists of a full-body Lycra equipped with 17 IMUs with magnetometers located at both shoulders, upper arms, lower arms, hands, upper legs, lower legs, feet, head, sternum, and pelvis. Data is wirelessly transmitted to a computer through Wi-Fi (IEEE 802.15.4). The accompanying software (Xsens MVN Studio version 4.2.1, Xsens Technologies B.V., Enschede, the Netherlands) allows to

Table 2.1: A description of trials in the experimental protocol (each trial was performed three times by all subjects). ADL = Activity of Daily Living, L = left and R = right.

	Trial	Short Description
Gait	1	Walk 10 m, walk 10 m, jog 10 m and sprint 10 m.
	2	Walk with a glass of water (dominant hand, non-dominant hand and in both hands)
	3	Walk 10 m, walk slowly 10 m, walk backwards 10 m, side-step six steps (L/R).
Sport	4	Two-legged jumps (4×), hops L/R (4×), run and jump L/R (2×), jump up (4×).
	5	Lunges L/R (4×), squats (4×), jumping jacks (4×).
	6	Sit-ups (5×) and side side-ups L/R (3×).
	7	Kick a ball against the wall L/R (3×).
	8	Throwing a ball against the wall L/R (3×).
	9	Crawling six steps.
ADL	10	Take a magazine, put it on the table, get seated, read a magazine, stand up and put it away.
	11	Take a tray with cups, walk with the tray, put it on the floor, stand up, pick it up.
	12	Take a glass, fill it with water and drink it in a chair.
	13	Put on a coat and take it off.
	14	Comb hair, scratch back, touch toes, rotate arms around shoulder back- and forward.
	15	Kneel down and tie shoelaces (L/R).
	16	Ascend and descend stairs.

visualize and export full-body (consisting of 23 body segments) kinematics of the subject at the selected sampling rate of 240 Hz. The inertial sensors used in Xsens MVN provide about 1 ° Root Mean Square Error (RMSE) in dynamic conditions and undisturbed magnetic field [214], whereas the joint angle accuracy of the system is in the range of 1 to 6 ° [217]. It should be noted, that sensor orientation may differ from segment orientation due to soft-tissue artifacts, which was 3 ° at largest for the knee joint angle [173].

A recording session starts with a simple calibration phase to estimate alignments between sensors and corresponding body segments. In this "N-pose" calibration, the subject is asked to stand upright, with arms next to body and palms facing the body, for a few seconds. For each tracking device, the system then fuses inertial and magnetic data to estimate sensor orientations which are further fed to another estimation engine that uses biomechanical constraints to estimate segment orientations and joint positions in a frame local to the laboratory (origin in the calibration position with x-axis aligned to magnetic North and z-axis upwards in the gravity direction) [161].

The proposed methods were validated using a reduced set of sensors, the orientations of considered individual sensors have been estimated using an Xsens tracking filter [168] and a calibration phase (N-pose) was performed as before to estimate corresponding segment orientations. Note that the use of a full-body motion capture system as an Xsens MVN allowed for the testing of different sensor configurations by just considering data from a subset of the 17 sensors.

2.2.4 Movement Database

All measured poses (about two million) define our movement database. Since movement data at 240 Hz resulted in almost indistinguishable adjacent poses, the movement database was down-sampled with a factor of 10, which resulted in approximately 200,000 poses. The size of the movement database is important as this

impacts training (for ANN) and testing (for NNS) times of learning algorithms. In order to avoid dependencies from global orientations of the body in the environment, all segment orientations were expressed with respect to that of a reference body segment. The pelvis was chosen as the reference segment, because it typically experiences lower dynamics and is therefore measured with higher accuracy. Furthermore, in many activities (e.g., ADL) there is no evident coupling between movements of the upper and lower extremities [56]. By splitting full-body poses into two separate databases (containing upper- and lower-body segments, respectively), a larger number of full-body poses can be accounted for (even though never explicitly performed by any subject) by combining instances from the two databases.

2.2.5 Learning Process

The learning process (both for NNS and ANN) aims at estimating a full-body pose starting from the five known individual segment orientations. As described in Section 2.2.4, the unknown upper- and lower-body segment orientations are independently estimated starting from known segment orientations of upper- and lower-body, respectively. The full-body pose is derived by simply combining upper- and lower-body estimates. Many different sensor configurations (varying number and placement of sensors) could be considered. For the sake of clarity, in Figure 2.1 we provide an intuitive visualization of the process in the case of five sensors being used on the pelvis, lower arms, and lower legs, respectively. In this case, pelvis and lower legs measured orientations are used to estimate the rest of the lower-body segments, whereas pelvis and lower arms ones are used to estimate the rest of the upper-body segments.

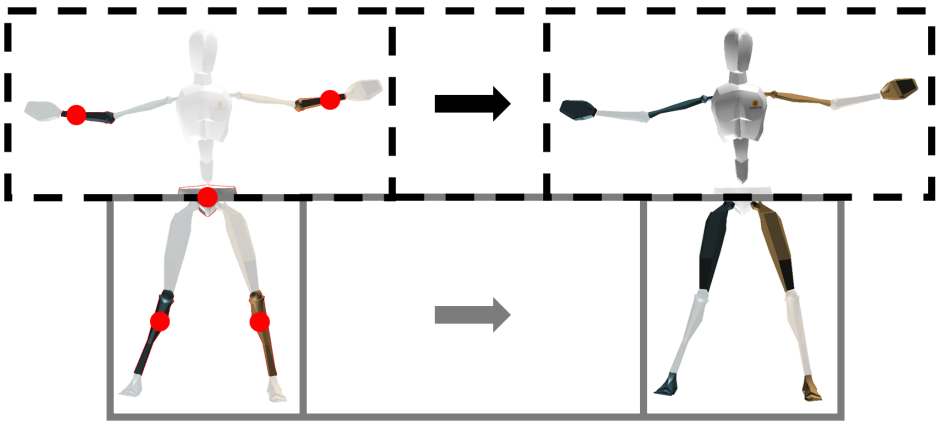


Figure 2.1: The left pose shows an example of input segment orientations (highlighted and sensor locations are marked with red dots) for both the upper- and lower-body estimations, whereas the right pose shows the corresponding output segment orientations (highlighted) for both the upper- and lower-body. Both estimations (displayed as black dotted and grey boxes) are combined to obtain the full-body pose.

The focus of this work is to investigate the performance of the two learning paradigms (lazy and eager) at estimating full-body poses from low dimensional orientation features. Therefore, we chose to evaluate the performance of both ANN and NNS in a "snapshot" manner, where poses are estimated independently of each

other, i.e., no temporal model that accounts for correlation between consecutive poses has been applied. Additionally, output of either algorithm was not explicitly corrected for poses implausible from a biomechanical viewpoint (e.g., knee flexion could result in angles larger than 180 °).

The implementation of ANN was designed using the neural network toolbox of MATLAB R2016a (Mathworks, Inc., Natick, MA, USA). Function fitting networks (`fitnet`) were chosen for their regression capabilities of estimating non-linear relations between inputs and outputs, as the mapping of a few segment orientations to a full-body pose is assumed to be non-linear. Weight and bias values of the ANNs were updated using the scaled conjugate gradient backpropagation(`trainscg`). The networks were trained for a maximum of 1000 epochs and the training was finished if the gradient stopped decreasing for 6 epochs (i.e., a local minimum of the regression problem was found). The inputs to the neural network were orientations (expressed as quaternions) of the measured body segments, and the outputs were chosen to be the orientations of the remaining body segments. Each input/output neuron processed a single element of a quaternion, where the norm of the output was not explicitly enforced to be one. The neural network was expected to learn this from the training dataset, however, the output was normalized to one, such that this was ensured.

The NNS implementation was similar to that of Giuberti [74]. The in- and outputs were identical to that of the ANN implementation. The distance (d_n) between measured orientation feature vectors and orientation feature vectors in the training dataset was computed using the mean quaternion shortest angle [89]:

$$d_n = \frac{\sum_{s=1}^{S} 2 \cdot \arccos\left([(q_s)^{-1} \otimes q_s^n]_1\right)}{S} \tag{2.1}$$

where q_s is the quaternion describing the orientation of the input body segment s (with a total of S input body segments), n is the index in the training dataset (with a total of N poses), $\otimes$ denotes the quaternion multiplication, $(\Delta)^{-1}$ is the quaternion inverse function and $[\Delta]_1$ extracts the first component of the vector. The computed distances (d_n) are used to compute a weighted pose, according to the following equation:

$$\tilde{p} = \sum_{l=1}^{k} (w_l \cdot p_l) \tag{2.2}$$

where p_l are poses in the training database, that are used to compute a weighted average full body pose ($\tilde{p}$) based on the k (neighbors) closest poses, where weights (w_n) are defined as:

$$w_n = \frac{max(\mathbf{d}) - d_n}{\sum_{n=1}^{N} (max(\mathbf{d}) - d_n)} \tag{2.3}$$

where $\mathbf{d}$ indicates a vector containing all calculated distances (d_n). The output orientations of pose $\tilde{p}$ were normalized to obtain proper unit quaternions, similar to the ANN implementation.

To avoid biased results, performance of both algorithms was tested independently on data from each subject (i.e., data from the same subject were never simultaneously appearing in training and testing). In particular, data from one subject was, in turn, used for testing, whereas data from the remaining five subjects was used for training.

Furthermore, an n-fold cross-validation [104] was performed to determine optimal algorithm parameters. Similarly to before, randomly splitting data into training and validation datasets might result in data from the same subject occurring in both datasets, thus introducing bias. Therefore, we chose $n = 5$ to allow us to split the training dataset subject-wise. Note that, since one subject is used, in turn, for testing, data from five subjects is left for training and cross-validation. The network configuration (for ANN) and the number of neighbors k (for NNS) were optimized in the cross-validation. For the sake of conciseness, the results of these cross-validations are presented in Table 2.2.

Table 2.2: Optimization settings for different learning algorithms. Network configuration describes the number of neurons in the first and second hidden layer. k = the number of neighbors, NNS = Nearest Neighbor Search and ANN = Artificial Neural Network.

Learning Algorithm	Optimization Parameters	Value
NNS	k	500
ANN	Network configuration	[250 100]

2.2.6 Performance Evaluation

Analyzing errors in the estimation of human motion is difficult, one has to deal with what people perceive as posed errors [107]. In similar works, either the Euclidean joint distance [74, 192] or the joint angle errors [36, 120] have been selected as error metrics. Furthermore, different applications may stress more errors on different parts of the body. A biomechanical application could focus more on specific joint angles, while a virtual reality application might be more interested in end-effector positions. For the sake of generality, both error metrics are reported and investigated in this work. In particular, joint angles are calculated as the relative orientation (to the proximal segment) of adjacent segments. The joint angle error (in three rotational directions) is then computed as the absolute difference between the measured and estimated joint angles. Joint positions are calculated using forward kinematics on segment lengths and orientations starting from the pelvis [206]. The joint position error is then computed as the Euclidean distance between the measured and estimated joint positions.

2.3 Results and Discussion

In this section, we evaluate the impact of sensor placement (Section 2.3.1) and different activities (Section 2.3.2) on the performance of NNS and ANN. Furthermore, generalization of such performance over different subjects (Section 2.3.3) and the impact of sensor noise (Section 2.3.4) are also investigated.

Accompanying videos can provide the reader with a clear intuition of algorithms performance. Nonetheless, for lack of space, in Figure 2.2 a representative selection of measured and estimated body poses for a few different activities (namely, walk, squatting, and kneeling) is shown.

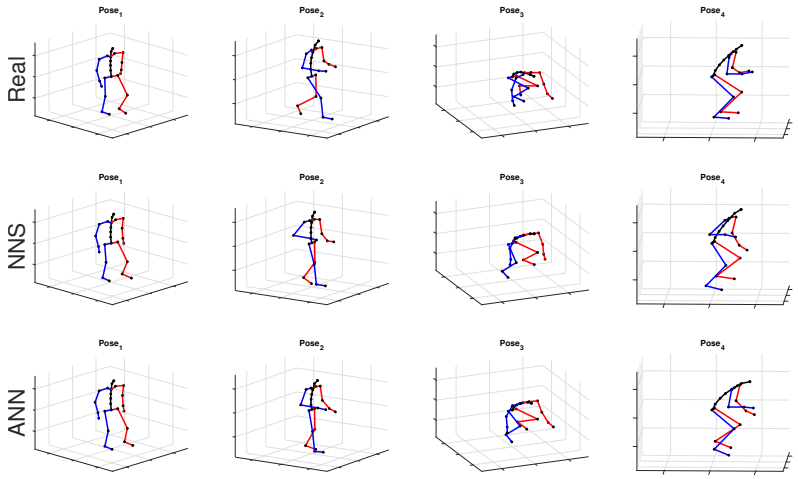
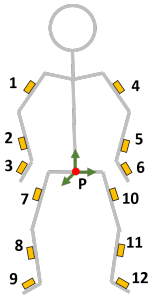


Figure 2.2: Measured and estimated poses of one subject are shown for different activities. Body segments on the left-side of the subject are colored red, while the right-side is colored blue. Joints are marked by black dots. The top row shows the measured poses of the testing dataset, the middle row shows poses estimated by NNS, and the bottom row shows those poses estimated by ANN (configuration D). Pose_1 shows a pose in mid swing, pose_2 is directly after toe off while carrying two cups, in pose_3, a tray of cups is picked up and the subject is squatting in pose_4.

2.3.1 Sensor Configuration Comparison



Configuration	Sensors Used
A	1, 4, 7 and 10
B	2, 5, 8 and 11
C	3, 6, 9 and 12
D	2, 4, 8 and 10
E	1, 2, 7 and 8
F	1, 2, 10 and 11

Figure 2.3: The different sensor configurations that have been investigated, each configuration is denoted by a letter (A–F) and the sensors for each of these configurations are numbered (1–12). (**left**) Sensor locations; (**right**) Sensors in the different configurations. In all configurations, the pelvis (P) is used as a reference segment.

Optimal sensor placement (on the body) in a reduced sensor system could be influenced by the requirements of the application of interest; in this work, we define it as the one that leads to the smallest average joint position/angle error. To limit the number of options, we investigated six configurations (all of them composed of five sensors), summarized in Figure 2.3. As mentioned in Section 2.2.4, the pelvis (shown as the red dot marked with P) is chosen as the reference segment for all configurations. Since, in our approach, upper- and lower-body pose estimation is separately performed, the chosen configurations have been defined so that the sensors are split uniformly between upper- and lower-body.

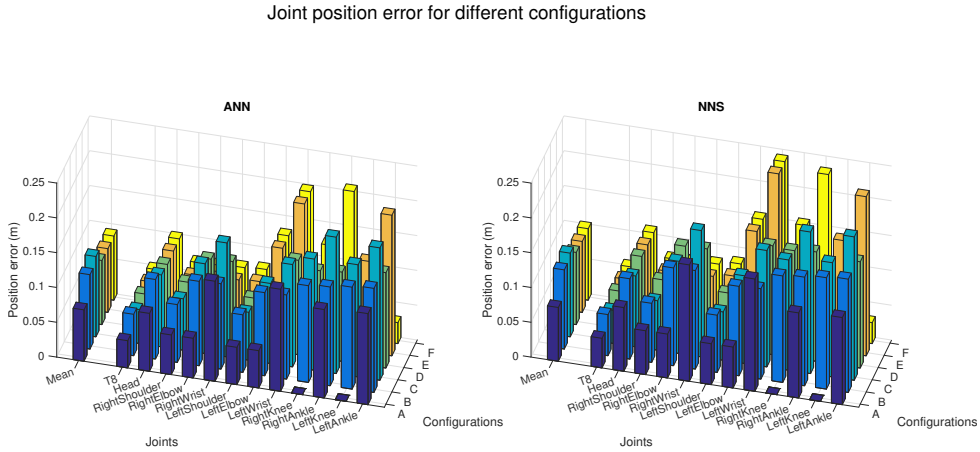


Figure 2.4: The left-side of both bar plots (ANN on the left and NNS on the right) shows the mean (over all six subjects) joint position error for the different configurations (as described in Figure 2.3). Individual mean (over all six subjects) joint position errors are shown on the right-side of both bar plots. A selection of joints is shown in both bar plots for readability. The different spine joints are not shown here because the joint position errors are comparable to the T8 joint. Additionally, the extra shoulder/foot joints are not presented because the magnitude of the error is similar to that of the shoulder and ankle joints shown.

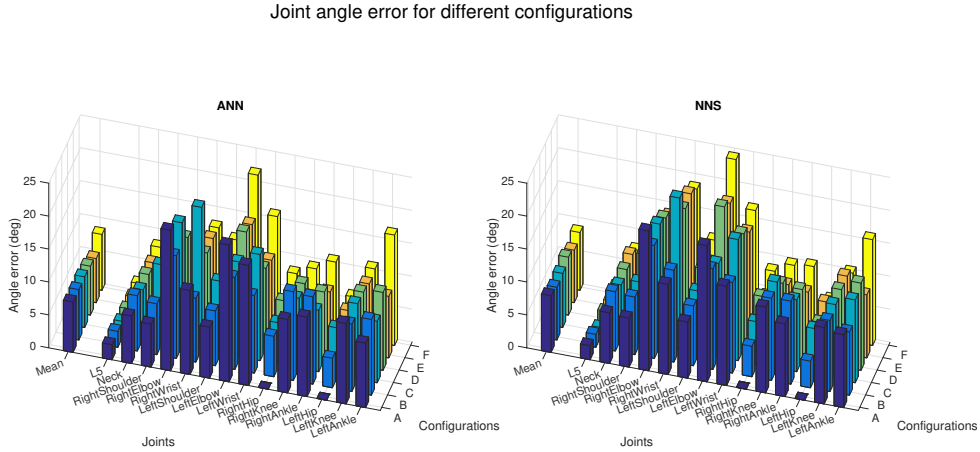


Figure 2.5: Mean (over all six subjects) joint angle errors for different configurations (as described in Figure 2.3) are shown on the left-side of both bar plots (ANN on the left and NNS on the right). Individual mean (over all six subjects) joint angle errors are shown on the right-side of both bar plots. A selection of joints is shown in both bar plots for readability. The different spine joints are not shown, because the joint angle errors are similar to the L5 joint. Similarly, extra shoulder/foot joints are not presented because the magnitude of the error is similar to that of the shoulder and ankle joints shown.

Figures 2.4 and 2.5 show the joint position error and the joint angle error (averaged over all six subjects), respectively, for the different sensor placements. Performance of ANN and NNS are reported, by also detailing contribution of each segment/joint (groups of bars on the right-side of plots) to the mean error (sin-

gle bars on the left-side of plots). In Figure 2.5, for the sake of conciseness, an average of joint angle errors, instead of errors in all three rotational directions (flexion/extension, abduction/adduction, and internal/external rotation), is shown. It can be observed that differences in performance between ANN and NNS are in the order of a few centimeters. The optimal sensor placement, according to the average joint position error, appears to be configuration A (namely, pelvis, upper legs, and upper arms) with 7 and 8 cm errors for ANN and NNS, respectively. These results are comparable to one of the reported situations in the work of Tautges et al. [192]. On the other hand, according to the average joint angle error, the optimal sensor placement is the one defined in configuration E (namely, an asymmetric configuration composed of pelvis, right upper and lower arm, right upper and lower leg) with errors equal to 7 and 8 ° for ANN and NNS, respectively.

Nonetheless, most of the considered configurations do not show major differences in average performance that would strongly motivate using one configuration over the remaining ones. Rather, it is important to highlight how errors on specific joints/segments are showing much larger variations. In that respect, it can be indeed noticed that average errors are mainly influenced by joint and segment errors which are particularly small due to the specific definition of the different configurations. For instance, configuration A shows shows hip angle errors of zero, because sensors are placed on adjacent body segments, consequently the knee joint position errors are zero.

In general, large individual joint position errors are shown at most distal joints (such as wrists and ankles) for all configurations, where errors could vary from 12 cm at the right ankle for configuration A (for NNS) to 24 cm at the left wrist for configurations E and F (for NNS). This is likely motivated by the fact that distal positions, since they are estimated using forward kinematics, tend to accumulate errors from segment to segment. On the other hand, joint angle errors show similar trends for different configurations. Even though these plots might better serve as tools for selecting optimal sensor placement in light of specific requirements of an application of interest, for the sake of a concise analysis, in the following we will investigate other aspects by choosing configuration D, which shows both mean joint angle and position errors close to the best found errors.

2.3.2 Activity Comparison

So far, no focus has been put on investigating evolution of error over time. Therefore, an example of the joint position error progression over time (using configuration D) for trial 1 (as defined in Table 2.1) of one of the subjects is shown in Figure 2.6. Different events/activities in the trial are marked by vertical dashed lines and labeled accordingly. Beside observing the quick and frequent jumps in the joint position error, which is likely the result of the use of a snapshot approach (i.e., no connection between adjacent poses), it is quite evident that different activities show different (yet consistent for each activity) error trends. This emphasizes the importance of characterizing the impact of different activities on the algorithm's performance.

By evaluating six different testing cases, as defined in Table 2.3, more insight into the performance of both ANN and NNS for different activities is provided. One trial for each activity was excluded from the training trials (of different subjects), such that capabilities of extrapolating movements could be analyzed.

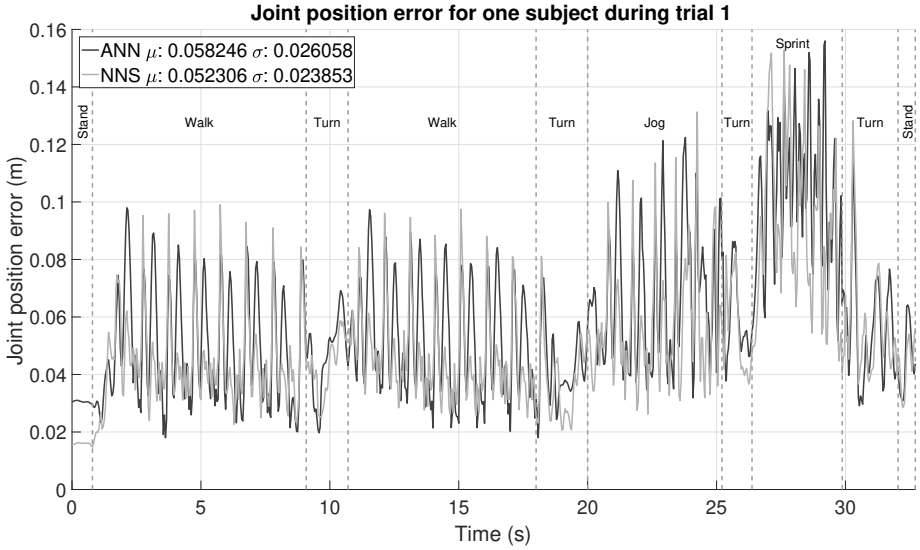


Figure 2.6: Average joint position error (using configuration D) of one subject during the first trial. Vertical dashed lines are shown to denote the different movements in that part of the trial.

Table 2.3: Description of the six testing cases, for the specified training database.

	Testing Case	Test trial	Training trials
Gait	A	1	1, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15 and 16
	B	2	
ADL	C	12	
	D	11	
Sport	E	4	
	F	5	

Figure 2.7 and 2.8 show the mean joint position error and mean joint angle error (averaged over all subjects), respectively, for the different testing cases. Note that, in Figure 2.8, the average flexion/extension angle error is shown, as this is the rotational direction with the largest variability for most joints and often of most interest for biomechanical analysis.

Performance differences between activity classes in mean joint position and angle errors are a few centimeters and degrees. This might be part explained by the fact that for different activities similar poses can occur (e.g., standing, walking, etc.), thus favoring classes with the largest overlap in poses (such as gait, where walking poses are indeed likely also occurring in ADL and Sport).

Trials that were excluded from the training dataset (testing cases B, D, and F) are estimated with similar accuracy (differences are in the range of a few centimeters and degrees) as the included trials (testing cases A, C, and E), which might indicate that generalizing between different motions within activity classes is possible. In general, there are not clear evidences that would favor the choice of one learning approach over the other. A larger difference between ANN and NNS can be observed for ADL, compared to Gait and Sport activities. This difference, although in line with the empirical standard deviation range, might be explained with the capability of

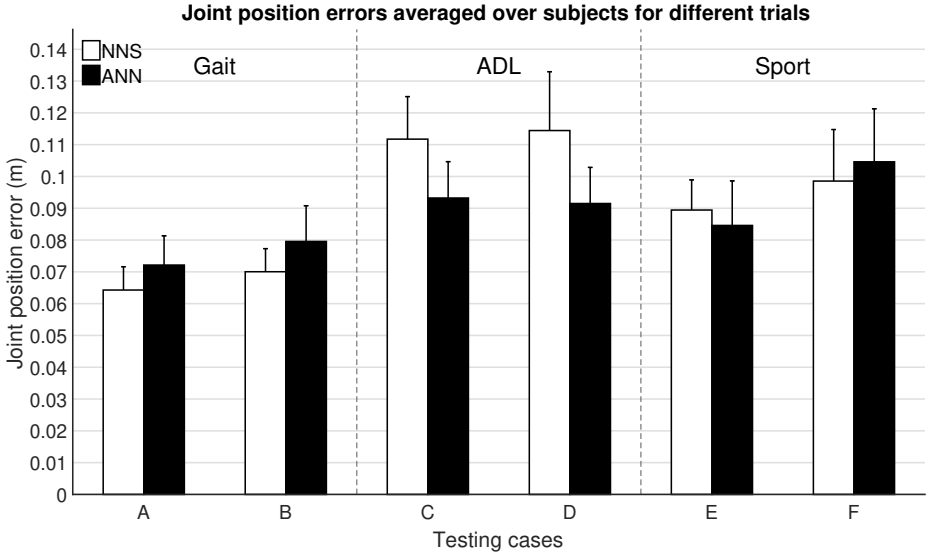


Figure 2.7: Mean (over all six subjects) and standard deviation (between the six subjects) of joint position errors for different testing cases, as described in Table 2.3.

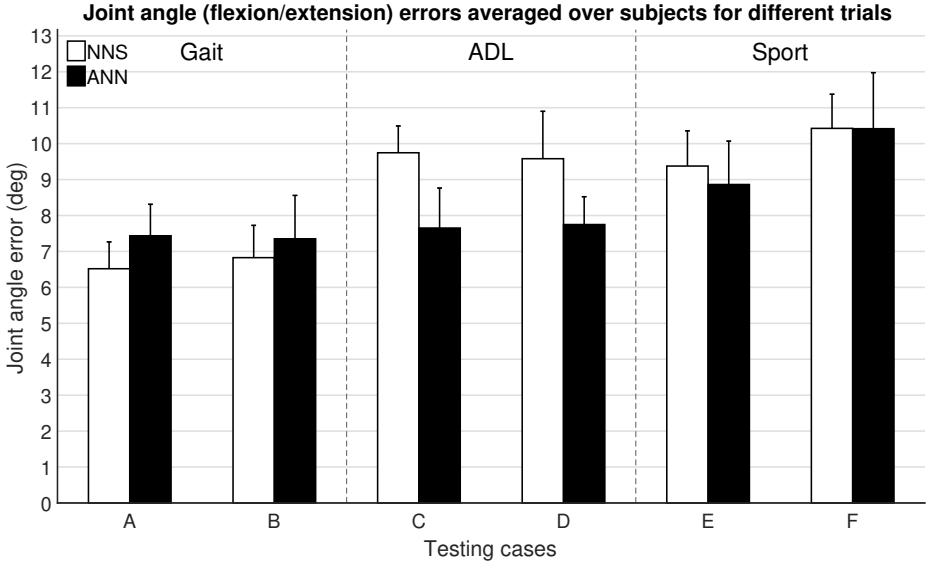


Figure 2.8: Mean (over all six subjects) and standard deviation (between the six subjects) of joint angle (flexion/extension) errors for different testing cases, as described in Table 2.3.

ANN of abstracting from the training database, while NNS depends on poses in the training database explicitly.

2.3.3 Generalizing Performance

Another aspect of interest is the difference in performance when testing over different subjects. In Figure 2.9, the distribution of the mean joint position error (over all trials) for each tested subject is shown, for both ANN and NNS. For ease of com-

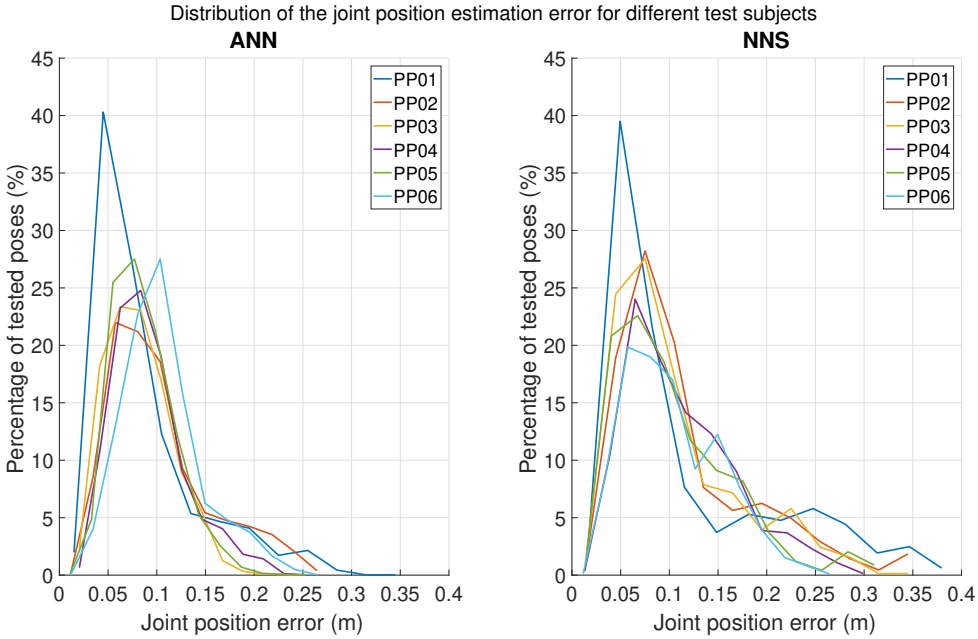


Figure 2.9: Distribution of joint position errors (with bins of 3 cm) for all scenario 1 activities for each subject. The left distribution shows ANN results, whereas NNS results are shown on the right.

parison, the distributions are overlapped and each tested subject is represented by a different color. The mode of the distributions for both ANN and NNS lies around 7 cm for all subjects and the whole distributions have similar shapes, indicating that the learning performances are quite generalizable over different subjects. Note that both ANN and NNS error distributions show long tails, which could be an indication of the (un)effectiveness of the algorithms at estimating rarely occurring poses in the movement database. Such error distribution profiles have also been reported by Tautges et al. [192].

2.3.4 Sensor Noise Analysis

Sensor orientation tracking from sensor fusion of IMU and magnetometers generally shows very small inclination errors and slightly larger heading errors, mostly due to the difficulty in determining magnetic North using magnetometers in the case of a magnetically disturbed environment. Although biomechanical body constraints can help in mitigating such errors if a full-body system is available, that may represent a harder challenge when using a reduced sensor set. During our data collection, we made sure that the environment was as clean as possible to guarantee the best quality in the collected data. Note that, this is even more important if such data are used as inputs for learning algorithms, as for our case. However, it is worth investigating the impact of noisy inputs on learning algorithms performance.

To illustrate the scale of such orientation errors in a typical real-world scenario, a measurement of about 2 h was performed using an Xsens motion tracker identical to the ones used for the data collection in this work. This measurement was

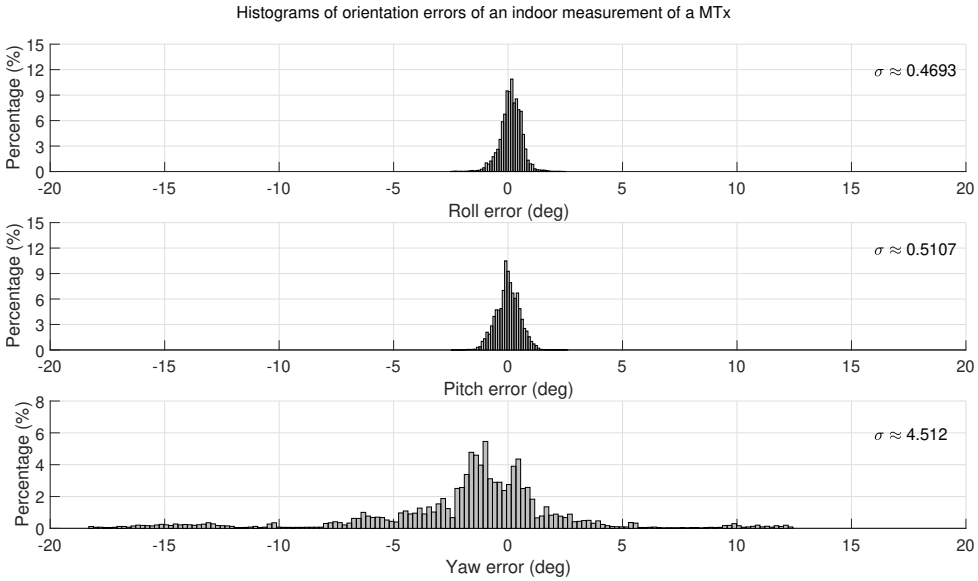


Figure 2.10: Orientational Root Mean Square Error (RMSE) errors of an Xsens motion tracker during a measurement of approximately 2 hours.

performed by a person carrying both sensors around in-/outside an office building (radiators, chairs, desks, and cars are examples of encountered sources of magnetic disturbances). A very accurate tactical grade IMU (0.75 °/h gyroscope drift, 1 mg accelerometer resolution) was used as a reference to estimate the error of the sensor orientation measured by the Xsens motion tracker. Histograms of the RMSE error of roll, pitch, and yaw, respectively, are shown in Figure 2.10. As expected, it can be observed that the yaw (heading) error is much larger ($\sigma \approx 4.51$) than that of the inclination (roll/pitch) ($\sigma \approx 0.47$, $\sigma \approx 0.51$).

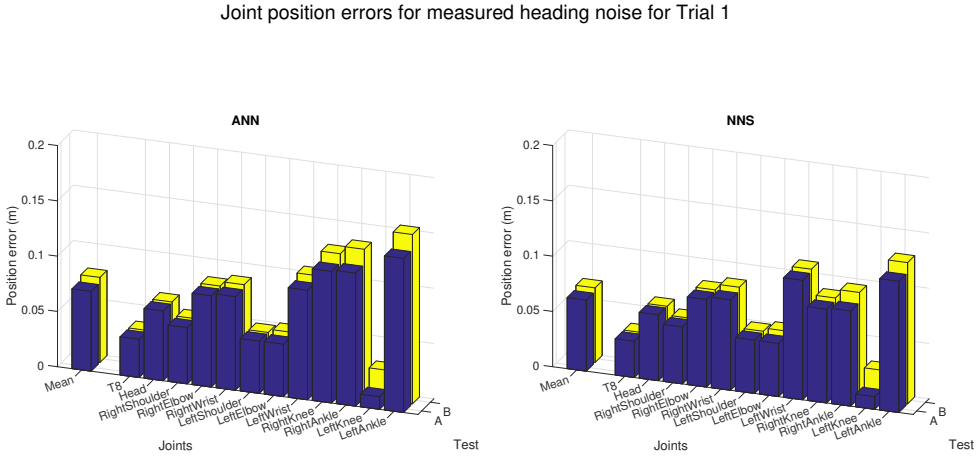


Figure 2.11: Mean (over all six subjects) joint position error obtained with ANN and NNS when the measured noise was applied to the heading orientation of the inputs. Here, A is the joint position errors obtained with the original input, whereas B is obtained with the measured heading errors applied to the original input.

Heading errors influence the measured orientation, however, its effects on the performance of learning algorithms remains unclear. Therefore, a white Euler rotation was applied to the measured sensor heading; the magnitude of this rotation was randomly drawn from the measured distribution (as shown in Figure 2.10). Note that, noise is overimposed only to testing data, since we do not want to corrupt the learned models. As our method uses a snapshot approach, a white signal is justified, since estimation errors are not a function of time. As expected, the joint position errors (averaged over all subjects) increase when noise is applied to the input (test B), as shown in Figure 2.11. Differences between tests A and B are in the range of 1 to 2 cm, whereas differences between both learning approaches are smaller than 1 centimeter.

Noise Sensitivity Analysis

In order to further investigate robustness of both learning approaches, a sensitivity analysis was performed on the errors in the heading direction. To that end, we assumed that the measured heading error could be approximated by a Gaussian distribution. Simulated heading errors (with varying standard deviations, shown on the x-axis) were applied to the measured heading orientations, for which the resulting mean joint position errors are presented in Figure 2.12.

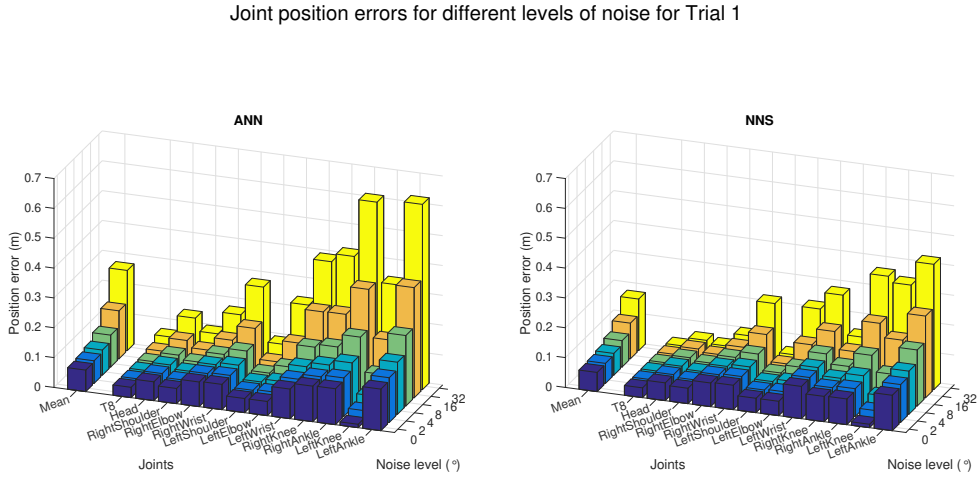


Figure 2.12: Mean (over all six subjects) joint position error obtained with ANN and NNS when a Gaussian white noise signal was applied to the heading orientation of the inputs. The magnitude of this rotation noise (standard deviation of the Gaussian white signal) is shown on the x-axis.

As expected, the mean joint position errors show an ascending trend for increasing levels of simulated noise. The 4 ° simulated heading noise results are similar to those obtained with the actually measured heading errors, confirming the validity of the Gaussian assumption. At this noise level, the mean joint position error is 7 cm for NNS, while it is 8 cm for ANN. This difference is largely the result of the position error increase at the ankles, which is 17 cm for ANN, and 14 cm for NNS. This can be explained by the fact that ANN uses a trained model, which was not trained for data with noise and could therefore provide implausible poses as output. The

estimated pose using NNS is more likely to be plausible, as it is a weighted average of poses in the training dataset. Therefore, NNS shows slightly smaller errors for increasing levels of noise on input orientations.

2.3.5 Computational Performance

In the introduction, (dis)advantages of the computational performance (training/ testing time and storage size) of lazy and eager learning approaches were mentioned. To provide more insight into this performance, these values were calculated for a training database of 124,214 poses and a testing database of 32,060 poses, of which results are shown in Table 2.4. In this example, the configuration of the ANN was two hidden layers with 250 neurons in the first layer and 100 neurons in the second layer. NNS exploited 500 neighbors in this implementation.

Learning Algorithm	Training Time (s)	Run Time (ms/sample)	Required Storage (MB)
ANN	525.9*	8.3	2.1
NNS	0	67.8	87.2

Table 2.4: Computation times (for a training database $N = 124,214$) for both ANN and NNS using single-core computation on a Core i7 @ 2.5 GHz system with MATLAB R2016a. Training time is the total required time, run time is presented as an average per sample, and required storage is the total size (as MATLAB variables stored as double) of the trained neural network or the training database. *ANN was trained using parallel (4 cores) computation on the same pc, as this computation can be performed offline.

As expected, a neural network estimates poses faster than NNS in the current implementation. The ANN results show potential for real-time estimation of full-body poses. An eager learning method, such as ANN, requires less storage space as the model is stored instead of the actual training data. As certain applications might require larger databases, the required storage might become an issue on a portable (embedded) system. Training time is less important if it is within reasonable boundaries, as training can be performed offline.

2.4 Conclusions

We have presented an in-depth performance analysis of ANN and NNS used for the estimation of full-body poses from orientations of a reduced set of IMUs (with magnetometers). The investigated approach showed a joint position error of approximately 8 cm and a joint angle error of approximately 7 °. The obtained results did not show clear evidence of an algorithm outperforming the other (differences in joint position and angle errors were shown to be approximately 1 cm and 2 °, respectively). Performance showed larger variations across different classes of activities, where smaller joint position/angle errors were obtained for gait, whereas ADL showed larger ones. Both algorithms have proved to be capable of generalizing over subjects. In a magnetic disturbed environment, NNS shows better performance (mean joint position error is 1 cm smaller) than ANN.

A choice for either algorithm would therefore depend on several factors, such as (but not limited to) computing power, real-time estimation, memory requirements, and/or magnetic disturbances. ANN is faster at run-time and requires less

memory, but training times can be long and performance is poorer when magnetic interferences are corrupting the input signals. NNS is flexible (no model creation), with better performance in magnetic environments, but memory requirements can be large, as well as computation times (especially for large databases).

2.5 Future Work

Jittering between consecutive poses has been observed (see, for instance, Figure 2.6 in Section 2.3.2) in the estimated pose outputs. In the approaches discussed in this work, this is most likely the consequence of not considering past poses in the estimation of the current poses. Other works have used priors in a Bayesian approach to ensure smoothness in the pose estimations [36, 152, 120, 192], which could be implemented in the current approach. An eager learning approach (such as a recurrent neural network) to predict a pose based on past poses could also be an interesting option. Additional information, such as biomechanical constraints, could be applied to improve pose estimation, as this would prevent implausible output poses, e.g., knee flexion angles cannot exceed 180 °. Supplementary features, based for instance on sensor accelerations, might further improve current pose estimates.

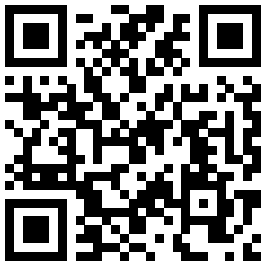
Finally, specific target application requirements might help to focus the development of a reduced sensor motion capture system, because the resulting errors can be evaluated within its context. Possible applications could be in virtual reality, sports, and/or in health care. The current implementation outputs orientations/positions relative to the body. However, a specific application, such as virtual reality, might require global motion as an output. To that end, global position tracking could be implemented using the current sensors in combination with contact detection.

2.6 Acknowledgments

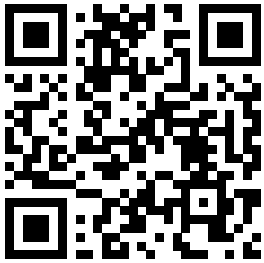
This research (project No.: 13917) is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs. The authors would like to thank the Roessingh Research & Development for the availability of the gait laboratory for the measurements.

2.7 Supplementary Material

A visualization of the motion capture dataset used in this work can be found at `https://youtu.be/v0xpWYlZVh0`:



A visualization of the results obtained from using either lazy or eager learning for estimating full-body poses using a minimal sensor set can be found at `https://youtu.be/zeUGTcb_8mI`:



Chapter

THREE

Time Coherent Full-Body Poses Estimated Using Only Five Inertial Sensors: Deep versus Shallow Learning

Published as:
F. J. Wouda, M. Giuberti, N. Rudigkeit, B. J. F. van Beijnum, M. Poel and P. H. Veltink **Time Coherent Full-Body Poses Estimated Using Only Five Inertial Sensors: Deep Versus Shallow Learning** Sensors 2019, 17, 3716 https://doi.org/10.3390/s19173716

Abstract

Full-body motion capture typically requires sensors/markers to be placed on each rigid body segment, which results in long setup times and is obtrusive. The number of sensors/markers can be reduced using deep learning or offline methods. However, this requires large training datasets and/or sufficient computational resources. Therefore, we investigate the following research question: "What is the performance of a shallow approach, compared to a deep learning one, for estimating time coherent full-body poses using only five inertial sensors?". We propose to incorporate past/future inertial sensor information into a stacked input vector, which is fed to a shallow neural network for estimating full-body poses. Shallow and deep learning approaches are compared using the same input vector configurations. Additionally, the inclusion of acceleration input is evaluated. The results show that a shallow learning approach can estimate full-body poses with a similar accuracy (∼6 cm) to that of a deep learning approach (∼7 cm). However, the jerk errors are smaller using the deep learning approach, which can be the effect of explicit recurrent modeling. Furthermore, it is shown that the delay using a shallow learning approach (72 ms) is smaller than that of a deep learning approach (117 ms).

Keywords: Inertial Motion Capture; Machine Learning; Neural Networks; Deep Learning; LSTM; Time Coherence; Human Movement; Reduced Sensor Set; Pose Estimation.

3.1 Introduction

Capturing full-body human motion can be valuable for various applications, such as biomechanical analysis, virtual/augmented reality, and gaming. For example, the increased use of wearable motion caption systems is helping coaches/athletes to improve their training programs [2]. Patients can benefit from biomechanical analyses to monitor treatment effectiveness [8]. Motion capture also has the potential to estimate kinetic quantities for various activities [7, 97, 209]. Virtual/augmented reality can produce realistic training environments for patients by providing interaction with the virtual elements using motion capture (e.g., knee osteoarthritis [99] or phantom limb pain [44]). The success of Microsoft Kinect shows that motion capture can also be applied to (serious) gaming (e.g., for traumatic brain injury patients [38] and neurological rehabilitation [103]). Full-body motion capture is currently done by using either body-worn sensors (e.g., inertial measurement units (IMUs) [170]) or external measurement equipment (e.g., cameras [146, 201]). These systems typically require users to wear sensors/markers on each (rigid) body segment, e.g., 17 sensors for Xsens MVN [214] and 37 markers for the Plug-In Gait protocol of Vicon [49, 201]. The (large) number of body-worn sensors/markers results in long setup times and can be obtrusive to the subjects.

Various studies have shown that using data-driven methods, a reduction in the number of sensors/markers for full-body motion capture is feasible by taking advantage of the inherent redundancy of human motion [165, 166, 195]. Chai and Hodgins have shown the potential of estimating full-body motion using only six retro-reflective markers, with a nearest neighbour search approach to map the low-dimensional marker input to full-body poses [36]. Note that poses are defined as the finite possible configurations of the body, i.e., a pose is a discrete sample of a motion sequence. However, their method, as most of the camera-based methods, is limited by the recording volume of the camera setup. To that end, Slyper and Hodgins used five body-worn accelerometers to estimate full-body motions in any environment, using a nearest neighbour approach as well [185]. The accelerometers were placed only on upper-body segments, which resulted in sub-optimal estimation performance of lower-body poses. This was further improved by Tautges et al. using a similar approach with four accelerometers placed on lower legs/arms [192]. These three methods include a cost function that weighs estimated poses in the past and present, which resulted in time coherent (plausible) output poses, since physically impossible large segment accelerations were smoothed.

All these approaches share the same "lazy learning" philosophy [3], since they don't learn a universal model to estimate full-body poses, but rather rely on a database of pre-recorded motion to look up at runtime. This approach is computationally demanding and often results in a delay between the performed motion and the estimated full-body pose. The significance of that delay depends on the application, e.g., virtual reality requires minimal delays as it can lead to motion sickness [75, 81], while providing feedback on gait analysis can be safely done with larger delays [99]. Opposed to lazy learning approaches, computation times can be reduced by using an eager learning approach (where a model is learnt and used at runtime), resulting in typically smaller delays. To that end, a shallow neural network was shown to estimate full-body poses from only five IMUs with comparable accuracy to lazy learning approaches [210]. However, estimated poses were not consistent over

time, since time relations were not explicitly taken into account. A short-term movement prediction can accurately be made given characteristics of the dynamic system at hand [213]. This has been shown by Von Marcard et al. with their optimization framework that uses data of six IMUs to estimate accurate time coherent full-body motion [202]. However, this approach cannot be applied in real-time, since a long data sequence is required for optimal performance.

Time coherency and real-time are both requirements for various applications that use full-body motions. Deep learning has the potential to provide time coherent real-time full-body pose estimates as shown by the increasing use in estimating human motions from video. For example, Fragkiadaki et al. used long short-term memory (LSTM) units in their recurrent neural network (RNN) architecture to estimate full-body kinematics from color videos [73]. Additionally, three-dimensional convolutional networks have been shown to be an effective network architecture for human activity recognition from videos [194]. Furthermore, optical motion capture can be complemented by inertial sensors and an LSTM architecture to improve visual tracking in the case of occlusions [148]. However, only bidirectional LSTM (bi-LSTM) units (to exploit information from both past and future) have been shown to accurately estimate full-body motion using the data of 6 IMUs. In this manner, time coherent (semi-)real-time output poses were achieved using an on-body measurement system. This approach was shown to result in the best performance by providing sequences that include past, current, and future frames as input; hence, output was delayed depending on the number of future frames required. However, such a deep learning architecture requires sufficient computational resources and a large dataset for training/evaluating.

In summary, estimating (real-time) full-body human motions from a minimal sensor setup can be achieved with good accuracy using offline methods (when a large motion sequence is available) or using deep learning at the expense of large datasets and computational resources. Our hypothesis is that similar results can be achieved by using a shallow learning approach. This resulted in the following research question: "What is the performance of a shallow approach, compared to a deep learning one, for estimating time coherent full-body poses using only five inertial sensors?" For the shallow learning approach, we propose a stacked input neural network (SINN) approach that requires smaller datasets and less computing power, which can result in suitability for real-time applications. This approach was based on earlier work of the authors [210], which showed good performance, but estimation of full-body pose at any given time only considered inputs at that instance, but not in the past. It therefore did not consider the inherent dynamics of the body that relate poses over time. In the current work, we developed a novel way of considering time dependencies in a shallow artificial neural network (ANN), namely, by moving complexity out of the (deep) network into a stacked input vector, which contains past and future information. The SINN approach was compared to a deep learning approach (with recurrent units) based on [88] (the current state-of-the-art for estimating full-body poses from a minimal set of inertial sensors), which is referred to as a recurrent neural network (RNN) for simplicity. It was chosen to use inertial sensors as input, since this allows for a wearable motion capture solution that does not require external infrastructure. Furthermore, it has been shown that differences in joint angles between optical and inertial motion capture are small [67, 188, 208]. To understand the performance of both the SINN and

RNN approaches, three aspects are analyzed in more detail, namely: configuration of the stacked input (e.g., number of past/future poses and time intervals), the inclusion of acceleration input information, and the computational cost (for training and evaluating).

3.2 Methods

3.2.1 Movement Dataset

The dataset contains a wide variety of movements performed by six subjects, as described in Table 3.1. Approximately 25 minutes of motion capture data was collected for each subject. Xsens MVN (Xsens Technologies B.V., Enschede, the Netherlands) was used for recording the subject's movements with 17 IMUs placed on (rigid) body segments at 240 Hz. Subjects performed a calibration pose to determine the sensor orientation with respect to the body, such that the biomechanical model of MVN Studio 4.2.1 (Xsens Technologies B.V., Enschede, the Netherlands) provides orientation of 23 body segments. The sampling frequency of 240 Hz resulted in nearly identical adjacent (in time) poses, and for most body parts, significant motion information lies well below 240 Hz. Therefore, the data were down-sampled by a factor of four (to 60 Hz), resulting in a dataset of approximately half million poses.

Table 3.1: A description of trials in the experimental protocol (each trial was performed three times by all six subjects). ADL = Activity of daily living, L = left and R = right [210].

	Trial	Short Description
Gait	1	Walk 10 m, walk 10 m, jog 10 m and sprint 10 m.
	2	Walk with a glass of water (dominant hand, non-dominant hand and in both hands)
	3	Walk 10 m, walk slowly 10 m, walk backwards 10 m, side-step six steps (L/R).
Sport	4	Lunges L/R (4×), squats (4×), jumping jacks (4×).
	5	Two-legged jumps (4×), hops L/R (4×), run and jump L/R (2×), jump up (4×).
	6	Sit-ups (5×) and side side-ups L/R (3×).
	7	Kick a ball against the wall L/R (3×).
	8	Throwing a ball against the wall L/R (3×).
	9	Crawling six steps.
ADL	10	Take a magazine, put it on the table, get seated, read a magazine, stand up and put it away.
	11	Take a tray with cups, walk with the tray, put it on the floor, stand up, pick it up.
	12	Take a glass, fill it with water and drink it in a chair.
	13	Put on a coat and take it off.
	14	Comb hair, scratch back, touch toes, rotate arms around shoulder back- and forward.
	15	Kneel down and tie shoelaces (L/R).
	16	Ascend and descend stairs.

3.2.2 Input Features

The recorded movement database contains orientations of 23 segments. Based on previous work in estimating full-body poses from a minimal body-worn sensor set [88, 192, 202, 209, 210], it was chosen to use the orientation of one segment for each limb as input features. Consistent with these works, the lower legs/arms and pelvis were selected as input segments (as highlighted in Figure 3.1), because they were positioned towards the ends of extremities. Orientation of these segments as

provided by Xsens MVN was used as input, while the remaining segment orientations were used for the corresponding output. The main reason for reducing the dimensionality is to take advantage of the fact that human body poses are extremely redundant if global body orientation is considered [165, 166, 195]. However, if orientations are expressed with respect to the body (e.g., pelvis), this dimensionality is further decreased. In this manner, the input/output space is reduced by relating all input/output orientations to the pelvis orientation (marked by the blue circle in Figure 3.1). The choice of the pelvis as a reference segment is motivated by its central location with respect to the different limbs.

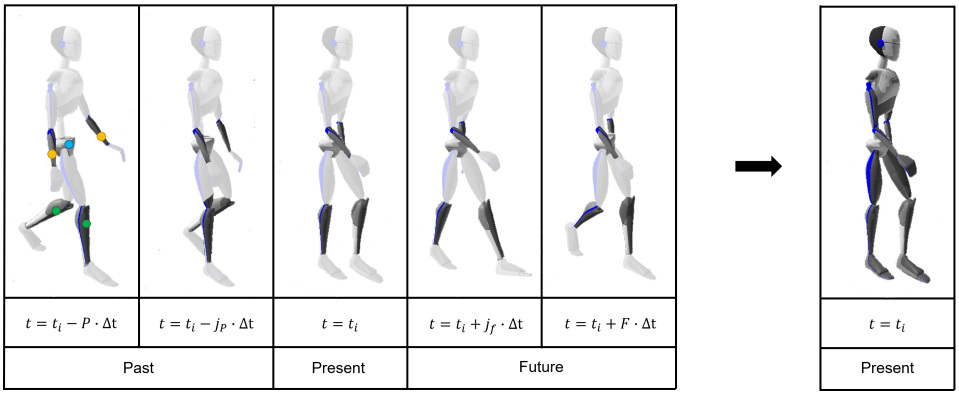


Figure 3.1: A sequence of inputs (lower arm (orange circle) /leg (green circle) orientations relative to the pelvis (blue circle) is used to estimate a single output pose (at time i). Size of the input sequence can vary by the number of past (P) and future (F) poses that are taken into account and the distance in time (Δt) between the different inputs. Here, j_p and j_f are used as counters for the past/future poses in time, which have a maximum value of P and F (in this example, $P = 2$ and $F = 2$) , respectively. In other words, $j_p = \{1, ..., P\}$ and $j_f = \{1, ..., F\}$, and Δt is defined as I/f_s with I as the sample interval and f_s as the sampling frequency.

It was chosen to train independent SINN and RNN to estimate upper- and lower-body poses using orientations (and accelerations) of two body segments. This was based on the limited dataset and difficulties with learning such complex relations for a shallow network, which was successfully applied in a previous work of the authors [210]. In other words, the lower arm orientations (relative to the pelvis, marked by the orange circles in Figure 3.1) are provided to a trained network to estimate the upper-body segment orientations (12 segments), and the lower leg orientations (relative to the pelvis, marked by the green circles in Figure 3.1) are input to a second trained network to estimate the lower-body segment orientations (6 segments).

Quaternions were used to represent orientations in the dataset (directly obtained from Xsens MVN), since this representation was shown to be fitting for training an ANN to map a reduced set of sensors to a full-body pose, after normalizing the output to obtain proper unit quaternions [89, 210]. Furthermore, in this manner, the input dimensions are smaller compared to rotation matrices, and they do not suffer from gimbal lock issues.

The IMUs measure 3D acceleration and angular velocity, of which the acceleration can be used as additional input to the network, since it provides information

about the linear movement and segment inclination, which could therefore result in better time coherency between output poses. This hypothesis was tested by training additional SINN and RNN (including acceleration features) and evaluating differences in performance compared to using no acceleration features. Accelerations are measured in the sensor frame, i.e., to compare different sensor accelerations, a transformation is required, as shown in Figure 3.2. The first step is rotating individual sensor accelerations to a common global frame, which is achieved using the orientation of those sensors (which are expressed in an identical global reference frame). The pelvis acceleration is then subtracted from the lower legs/arms such that a relative acceleration is obtained, which also removes the gravitational acceleration from the resulting relative measure. Rotating that outcome to the pelvis orientation results in acceleration features that are relative to the reference segment and are not affected by the orientation of the body w.r.t. the world.

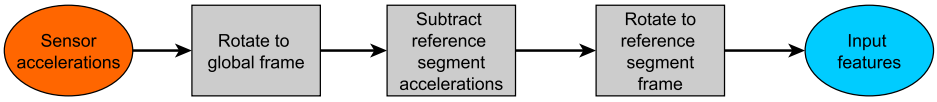


Figure 3.2: Processing of the measured sensor accelerations to be suitable input to the recurrent neural network (RNN) and stacked input neural network (SINN).

3.2.3 Stacked Inputs

The SINN was trained to map a sequence of inputs to one single full-body pose (i.e., input $x = \begin{bmatrix} t_i - P \cdot \Delta t & t_i - j_P \cdot \Delta t & t_i & t_i + j_f \cdot \Delta t & t_i + F \cdot \Delta t \end{bmatrix}$), as depicted in Figure 3.1. Δt is the time between different poses, which is defined as $\Delta t = I/f_s$, where I defines sample interval and f_s the sample frequency (60 Hz). j_P is a counter for the past poses (P), while j_F is a counter for the future (F) poses that are taken into account. Additionally, acceleration features can be appended to this input matrix if required.

By doing this, we want to prove that the proposed SINN is able to "learn" time coherency even when past and future sensor data are stacked into the same input vector. This approach allows for various options for the number of poses over time that are considered ($SIL = P + 1 + F$, stacked input length) and the sample interval (I) between those adjacent poses. The optimal configuration depends on the requirements of the application, e.g., real-time applications, required accuracy and movement types. Figure 3.1 shows an example of a gait sequence using $P = 2$, $F = 2$ and $I = 8$ (with $\Delta t = 1/60$ s); therefore the length of the shown sequence is 5 samples that span approximately 0.55 seconds.

3.2.4 Network Architecture

Figure 3.3 depicts the network architectures for the deep (RNN) and shallow (SINN) learning approaches. The implemented RNN was inspired by the work of Huang et al. using bi-LSTM layers [83, 88, 172]. The network architecture of the RNN allows for recurrency and hence no input stacking is required. However, due to the bidirectional units in the networks, better qualitative and quantitative results can be obtained

by processing a sequence of inputs (which is implemented as a sliding window) [88]. Therefore, both the RNN and SINN will be trained/evaluated using identical input sequences (in both length and configuration) obtained from our collected dataset. The difference is that the RNN gets a matrix of size (8,SIL) as input, while this is stacked to (8·SIL,1) for the SINN.

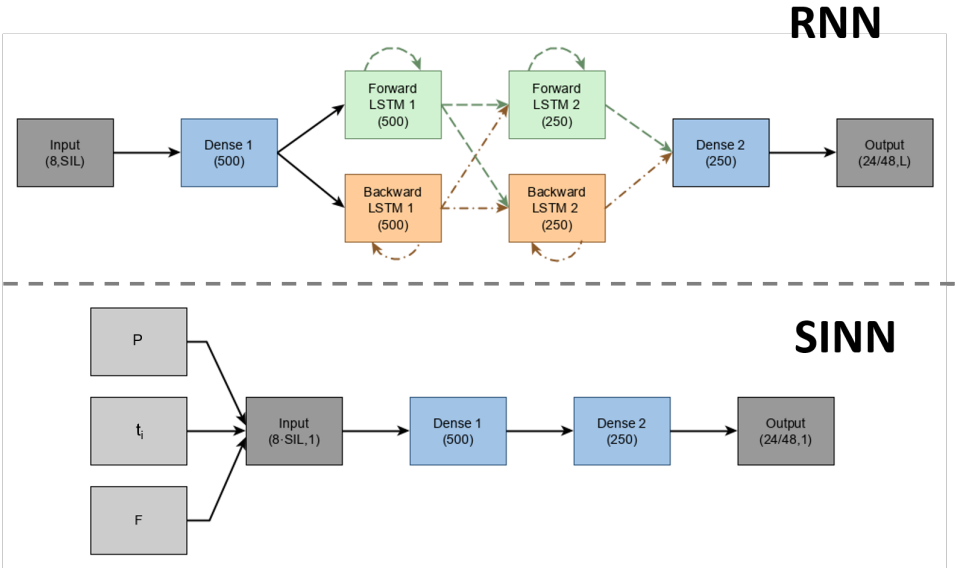


Figure 3.3: The implemented network architectures for the deep (RNN) and shallow (SINN) learning approaches. Different networks were trained for estimating upper/lower-body poses, which resulted in 8 inputs (2 segment orientations, represented by quaternions, relative to the pelvis) times SIL (stacked input length) poses for both SINN and RNN. Furthermore, a different number of outputs was obtained from the separate networks, namely, 24 for the lower body and 48 for the upper body . Input to the SINN is stacked with adjacent poses from past (P), current (t_i) and future (F) time samples, resulting in a total of L samples that are taken into account. The same sequence can be provided as an input matrix to the RNN, which produces a sequence as output (and the relevant pose can be used). The different types of hidden layers are shown by the various colors with the corresponding number of neurons shown in brackets.

MATLAB R2018b (MathWorks, Inc., Natick, MA, USA) was used to implement both the SINN and RNN approaches. The following training parameters for the bi-LSTM network were the same as in [88], namely, using an Adam optimizer [102], identical learning rates and dropout [88]. The number of neurons per layer and number of hidden layers was chosen based on a previous work of the authors and was validated by comparing various network sizes. The number of hidden layers for the RNN approach was based on the work of Huang et al. [88]. Separate RNNs were trained for the upper- and lower-body movements due to the limited dataset and to provide a fair comparison with the proposed shallow approach (SINN).

Both the RNN and SINN were trained using a subject-wise six-fold cross-evaluation (5 subjects for training and 1 for evaluating). This evaluation approach provides information about the generalization performance of the trained networks over the subjects, as different subjects are in the training/test set for each evaluation [22, 104].

3.2.5 Performance Evaluation

As commonly done in similar works, the accuracy of full-body poses is evaluated using either Euclidean joint distance [73, 210] or joint angle errors [36, 202]. The application largely determines the evaluation metric of interest. For conciseness, joint position errors are reported in the current paper, as these were shown to be related to how pose similarity is perceived by humans [107]. This was evaluated by calculating the Euclidean norm between the full-body joint positions obtained from Xsens MVN and the estimated joint positions from the SINN/RNN approaches. These 23 joint position errors were then averaged to obtain a mean error value for each pose.

Furthermore, a previous work of the authors has shown that jitter was present in the outcome pose sequences. It was chosen to quantify jitter by calculating jerk (the third derivative of the joint positions), since this provides insight in the smoothness error [71]. This was evaluated by calculating the Euclidean norm between the ground truth (Xsens MVN) and estimated jerk (SINN/RNN approaches).

Real-time applications require a delay that is not larger than the threshold that results in motion sickness [75, 81]. However, this threshold is individual: a delay of 100 ms might be acceptable for some people, while others might not cope well with delays larger than 20 ms in virtual reality applications [91]. Therefore, the processing time of both the SINN and RNN approaches was evaluated on a notebook (Lenovo ThinkPad W540, Beijing, China) (CPU i7-4710MQ @ 2.50 GHz, 8 GB RAM, NVIDIA Quadro K1100M, Santa Clara, CA, USA), which is representative of equipment that can be used for a real-time application. As the number of future poses impact the additional delay between the measured movements and the estimated full-body poses, this parameter will be regarded. Furthermore, the difference in training time between both approaches was evaluated on a high-end machine (equipped with one NVIDIA GTX Titan X (Pascal 12 GB)), since training requires more computational resources.

3.3 Results

In this section, we explore the impact of different time window configurations on the SINN/RNN performance (Section 3.3.1). Furthermore, the addition of accelerometer data in the input is investigated (Section 3.3.2) and the computational cost of both approaches is compared (Section 3.3.3). Additionally, videos of the obtained output have been included as supplementary material.

3.3.1 Time Window Configurations

The considered dataset consists of three activity types, namely, gait, sports and ADL. The dynamics of different activities can vary substantially, which could result in variation of the optimal time windows (length, configuration and spacing). To that end, mean (over six subjects) joint position and jerk errors are presented for a representative trial within each of those activity classes.

Figure 3.4A shows the mean joint position errors for a gait trial, namely, for trial 1 as described in Table 3.1. With increasing distance (I) between included samples, the mean joint position error shows an increase for all different SIL with the RNN

approach. This trend can also be seen for the SINN approach; however, the absolute error increase is smaller. Furthermore, the standard deviation of all mean joint position errors (for both approaches) is of similar small size (∼0.01 m), indicating good generalization over different subjects. The impact of using information from past or future (stacked input configuration) is shown to be minimal, while the SIL (number of samples) has a larger effect.

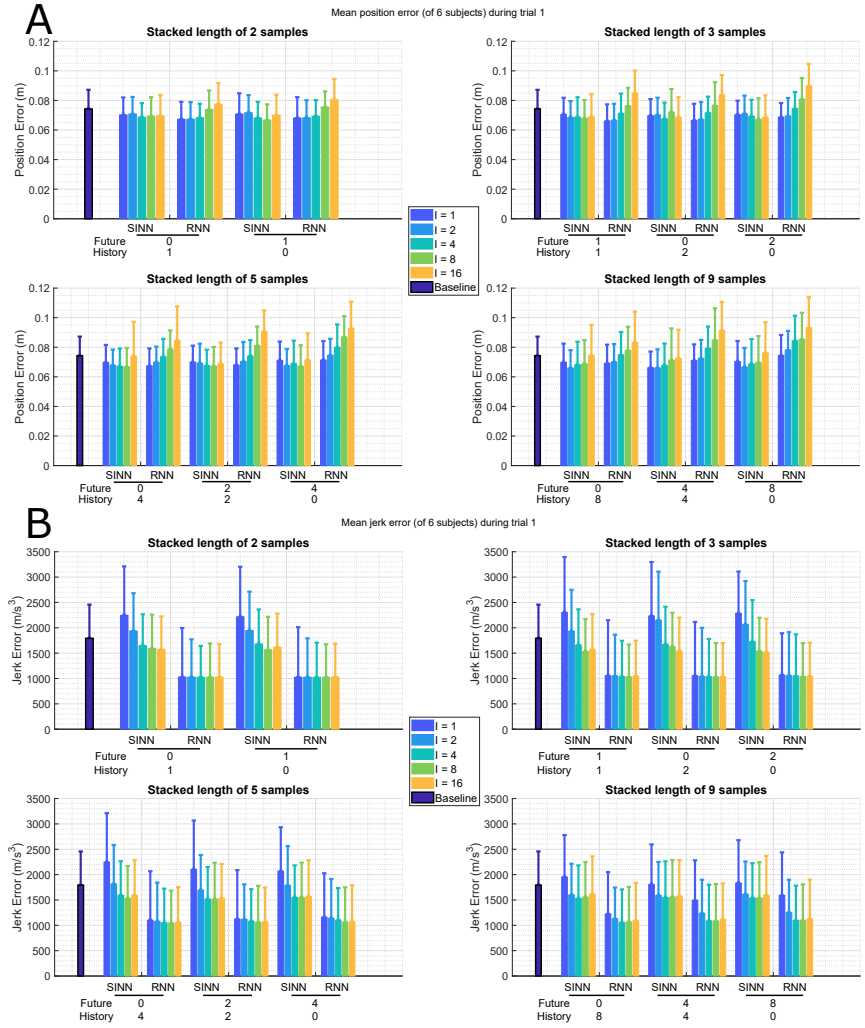


Figure 3.4: Bar plots of the full-body mean (of 6 subjects) joint position (**A**) and jerk (**B**) error for the shallow (SINN, left bars for each configuration) and deep (RNN, right bars for each configuration) learning approaches during a gait trial (1 in Table 3.1), standard deviation over the various subjects is displayed by whiskers. The different time windows are presented on the x-axis, where the number of past (P) and future (F) poses are shown. The interval (I) between input poses are marked by the different colors, where the number of samples between input poses is shown. For comparison, the mean joint position error (**A**) for using only the current pose as input (SINN approach) is 0.07 ($\pm$0.01) m. For comparison, the baseline using only the current pose as input) mean joint position (**A**) and jerk (**B**) errors are shown as the dark blue bars on the left.

Figure 3.4B shows the mean (and standard deviation) joint jerk errors for the different time window configurations of trial 1, as described in Table 3.1. The RNN shows smaller jerk errors than the SINN approach, which was to be expected since the RNN explicitly takes into account time coherency of the different poses in the input sequence. Furthermore, for increasing intervals (I), smaller joint jerk errors can be seen in the SINN approach, while this effect is only shown for the larger SIL (of 5/9 samples) of the RNN approach. Compared to using only a single pose as input to the SINN, a decrease in joint jerk error is observed when the SIL is larger (i.e., more time information is taken into account). Furthermore, the stacked input configuration has a smaller effect on performance than the SIL, similar to what was observed for the joint position errors.

Overall increase in joint position errors for sports activities compared to gait is observed for both approaches (**SINN**: $median(gait) = 0.069$ m and $median(sports) = 0.079$ m; **RNN**: $median(gait) = 0.076$ m and $median(sports) = 0.089$ m) in a more dynamic activity, as can be seen for trial 5 in Figure 3.5A, which includes sport related tasks (as described in Table 3.1). For the SINN approach, the interval (I) shows a smaller effect on the joint position error than for the RNN approach. A decrease in the joint position errors for both approaches can be seen when more information is taken into account (larger SIL).

As was to be expected for a more dynamic trial, the joint jerk error is larger for sports activities than for a gait trial (**SINN**: $median(gait) = 1609$ m/s^3 and $median(sports) = 2309$ m/s^3; **RNN**: $median(gait) = 1056$ m/s^3 and $median(sports) = 1515$ m/s^3), shown in Figure 3.5B. Similar to the gait trial, a decrease in joint jerk error is observed for large I; however, this effect decreases for larger SILs. Furthermore, both approaches show minimal differences in performance for various stacked input configurations (P/F) with a fixed SIL.

Figure 3.6A,B shows, respectively, the joint position and jerk errors during an ADL (as described in Table 3.1). Similar to the previous two activities, an error decrease is observed for larger SIL, but no large differences for the various configurations (P/F) are observed in this activity either. A decrease in performance is shown for larger I but is not consistent between both approaches.

Even though the dynamics are different between the three activities, it can be seen that small I results in a joint position error increase (shown in Figures 3.4A, 3.5A, and 3.6A), as the similarity between those poses is too high, and therefore, the individual poses contain minimal additional information, while large I results in an increase of the joint jerk errors (shown in Figures 3.4B, 3.5B, and 3.6B), since dependency between poses decreases at larger time intervals. In other words, there is an optimal interval I, which depends on the specific dynamic nature of the activity. This optimum can therefore be found around $I = 2$ and $I = 4$, since this results in smaller joint position and jerk errors on average over the various subjects and activities.

Configuration of the input vector (past/future) shows smaller effects on the joint position/jerk errors than the SIL. This effectively means that a longer sequence of inputs is more beneficial than changing the configuration of those inputs, e.g., including future information at the expense of past information. However, a marginal error decrease was observed when future information was included. Hence, for real-time applications a larger SIL can be sufficient, while for applications that require higher accuracy, it can be valuable to include future information.

For sake of simplicity, in the remainder of the paper, we will use $I = 2$, $P = 2$ and $F = 2$ as this configuration resulted in an acceptable trade-off of accuracy and possibilities for real-time applications.

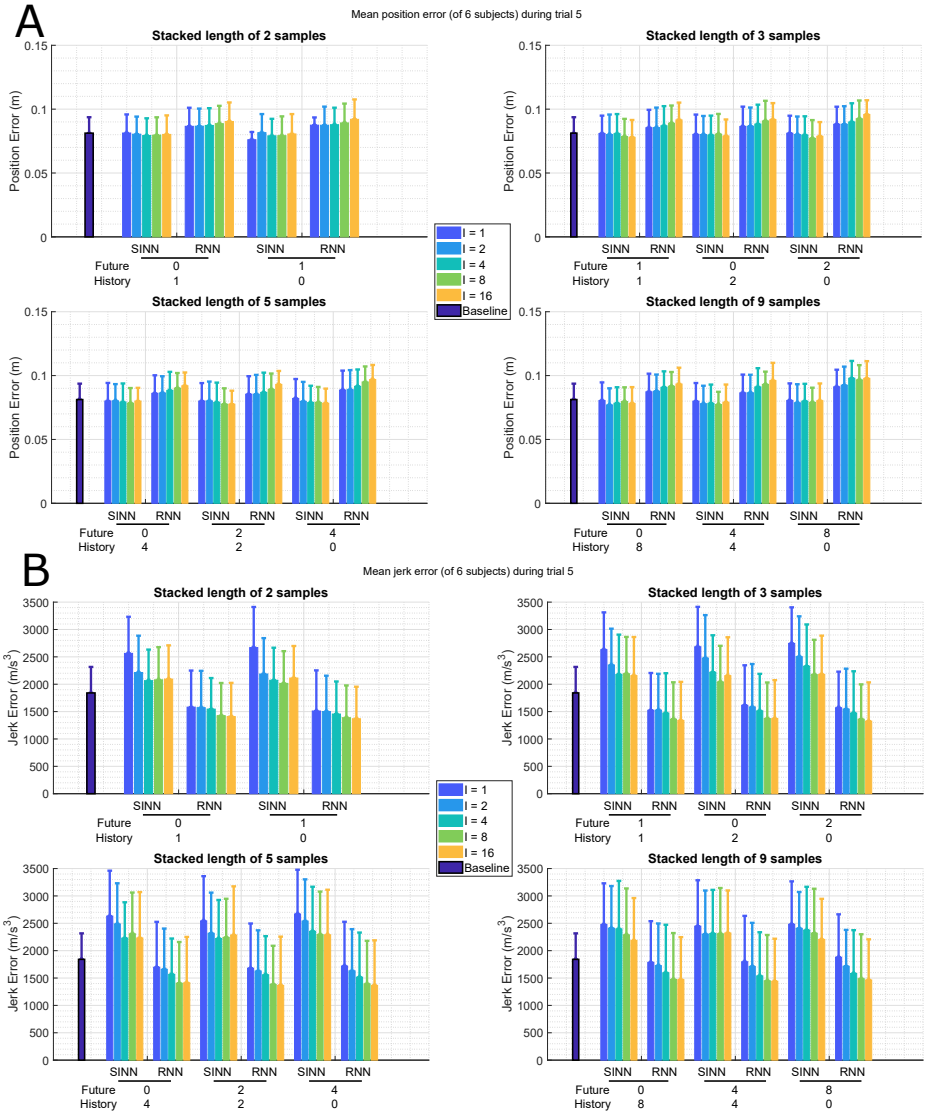


Figure 3.5: Bar plots of the full-body mean (of 6 subjects) joint position (**A**) and jerk (**B**) error for the shallow (SINN, left bars for each configuration) and deep (RNN, right bars for each configuration) learning approaches during an ADL trial (5 in Table 3.1), standard deviation over the various subjects is displayed by whiskers. The different time windows are presented on the x-axis, where the number of past (P) and future (F) poses are shown. The interval (I) between input poses are marked by the different colors, where the number of samples between input poses is shown. For comparison, the mean joint position error (**A**) for using only the current pose as input (SINN approach) is 0.08 ($\pm$0.01) m. For comparison, the mean joint jerk error (**B**) for using only the current pose as input (SINN approach) is 1.8 ($\pm$0.5) $\times 10^3$ m/s^3.

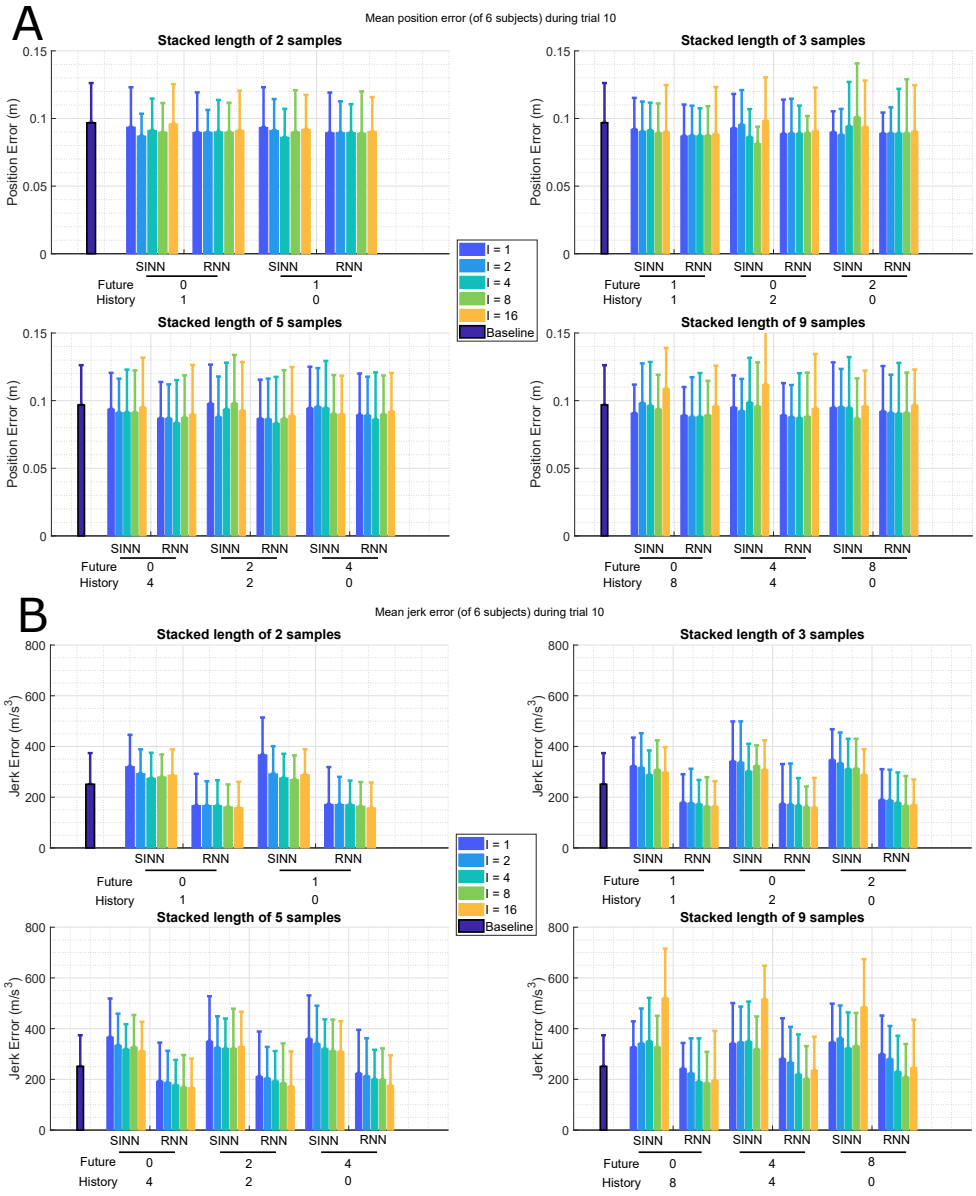


Figure 3.6: Bar plots of the full-body mean (of 6 subjects) joint position (**A**) and jerk (**B**) error for the shallow (SINN, left bars for each configuration) and deep (RNN, right bars for each configuration) learning approaches during a sports trial (10 in Table 3.1), standard deviation over the various subjects is displayed by whiskers. The different time windows are presented on the x-axis, where the number of past (P) and future (F) poses are shown. The interval (I) between input poses are marked by the different colors, where the number of samples between input poses is shown. For comparison, the baseline (L**A**) for using only the current pose as input mean joint position (**A**) and jerk (**B**) errors are shown as the dark blue bars on the left.

3.3.2 Including Sensor Acceleration Features

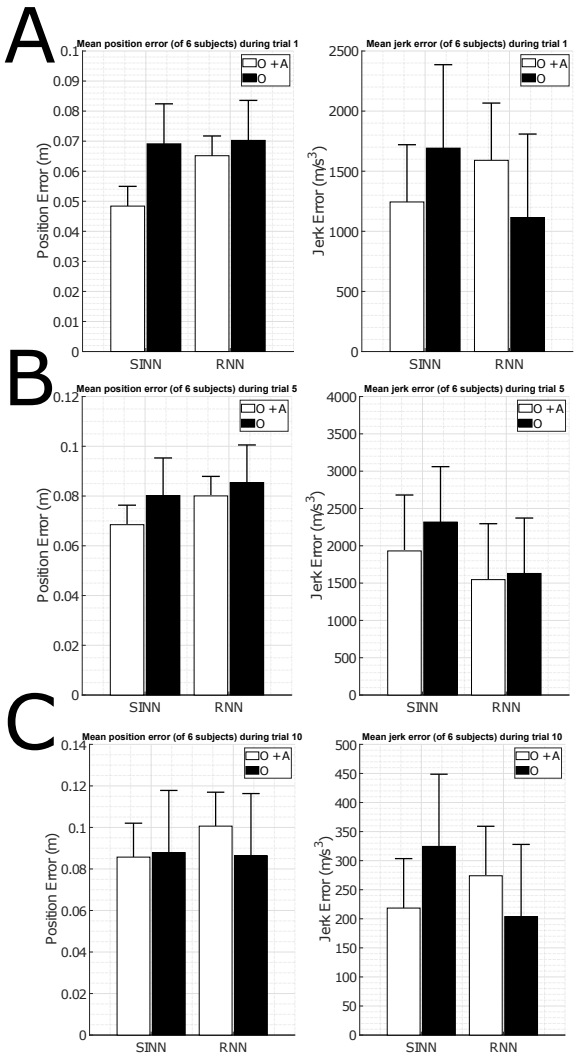


Figure 3.7: Bar plots of the mean (of 6 subjects) joint position and jerk error for the shallow (SINN) and deep (RNN) learning approaches (left and right, respectively) using orientation features (**O** in black) and including accelerations (**O** + **A** in white). Three different types of activities are shown namely: gait (**A**), sports (**B**) and ADL (**C**). Standard deviation over the various subjects is displayed by black whiskers. These results were obtained using the following parameters: $I = 2$, $H = 2$ and $F = 2$.

Figure 3.7 shows the mean joint position/jerk errors for three different activities, namely, gait, sports and ADL, using orientation features (as shown in Section 3.3.1) and including accelerations. The gait and sport trials show a decrease in joint position error for both the SINN and RNN approaches when acceleration features are included compared to only orientation features. This is not the case for the ADL trial, which could possibly be the result of decreased dynamics in ADL tasks, hence

acceleration information could provide less additional knowledge. The joint jerk errors are smaller when acceleration features are included for the SINN approach. This is only observed for the sports trial in the RNN approach. This indicates that including acceleration features improves full-body pose estimation, but it can be at the expense of smoothness of the output.

3.3.3 Delay Assessment

A delay is already introduced by using future sensor information as input to both the SINN and RNN approaches. The chosen input sequence configuration (2 future poses) results in a delay of 67 ms using MATLAB R2018b (MathWorks, Inc., Natick, MA, USA). In Table 3.2, the training and calculation times are presented. It should be noted that these results are obtained from training/evaluating a SINN and RNN for both the upper and lower body, which requires double the amount of training time, but could be performed in parallel. It can be seen that a shallow network, as expected, is faster to train and evaluate. However, the RNN approach does not require a new training cycle when experimenting with the input sequence configuration, e.g., when more past/future frames should be taken into account.

Table 3.2: Training of both approaches was done using a machine equipped with a single NVIDIA GTX Titan X (Pascal 12 GB) with MATLAB R2018b. Testing was done on a notebook, namely, a Lenovo ThinkPad W540 (CPU i7-4710MQ @ 2.50 GHz, 8 GB RAM, NVIDIA Quadro K1100M) with MATLAB R2018b.

Approach	Training Time (Hours)	Evaluation (ms/Sample)
RNN	∼6	∼50
SINN	∼1	∼5

3.4 Discussion

In this work, we have shown that using either a shallow (SINN) or deep (RNN) learning approach for estimating full-body poses using only five IMUs placed on the lower legs/arms and the pelvis results in similarly accurate outcomes.

A limitation of this work is that input to both learning approaches was the segment orientation from Xsens MVN, i.e., the calibrated sensor to segment pose data were based on the full-body approach [170]. This full-body approach benefits from assumptions based on dynamics of a human body, which is not the case for a single sensor. However, the orientation accuracy of an IMU is within 0.5 degrees [119]. Furthermore, an application with only five inertial sensors would require a sensor to segment calibration, such as the static neutral pose proposed by Huang et al [88]. A misalignment of the sensors with respect to the calibration pose could decrease the accuracy of the estimated full-body poses; however, it has been shown that sensor noise has a minimal impact on the performance of such a trained neural network [210]. The SINN/RNN approaches could also be trained to handle inputs with noise, such as was done for optical pose estimation [180].

The input and reference data consisted of inertial motion capture data. In this manner, no external infrastructure is required for data collection. The accuracy

of inertial motion capture is comparable to that of optical systems when looking at joint angles [67, 188, 208]. However, the use of optical motion capture data provides interesting opportunities for enlarging the training dataset, since such data are publicly available [33, 122, 129]. Our proposed SINN approach could benefit from such datasets if they were recorded with sufficient optical markers, since it requires three-dimensional orientations as in/output. The relatively small dataset used in this work allowed for a fair comparison with the previous work of the authors [210] and demonstrated that a shallow approach trained on such a dataset can estimate full-body poses with good accuracy. It should be noted that a shallow network requires significantly fewer parameters to be trained (the SINN approach has approximately 25 times fewer trained parameters compared to the RNN approach), which impacts the minimal required training dataset size.

It was chosen to use $I = 2$, $P = 2$ and $F = 2$ as an input configuration to evaluate the addition of acceleration features, due to it being an acceptable trade-off between accuracy and delay. However, this choice largely depends on the application and dataset as differences were also observed for the various activities. Variations in the optimal settings for the various activities could be the effect of differences in the involved dynamics. However, more insight in this relation is required, which could reduce the search for optimal settings for specific applications. Therefore, this setting is not a final recommendation, but the presented results can provide a direction for specific applications.

The gait trial was estimated with the smallest position error, which was to be expected due to the repetitive and cyclic nature of the activity. The largest errors can be found in the activities that are less cyclical, such as ADL. A trend observed for all activities is that the position error decreases when more information was used as input (SIL). This effect was larger than changing the input vector configuration (number of poses from past/future). However, including information about the future resulted in a decrease of joint jerk errors, i.e., more smooth outcomes. This was to be expected, since interpolating is a less error-prone task than extrapolating [84]. Joint jerk errors were further improved by including acceleration information as input to both the RNN and SINN approaches, as can be seen from the improved joint position errors for all trials.

These observations of use of a RNN approach for estimating full-body movements using a minimal sensor set are consistent with the findings presented in [88]. Since a different dataset for training/testing was used in their work, which could indicate that these effects are not dataset-dependent. Joint position accuracy reported by Huang et al. was 6.49 cm on average (for their RNN approach) compared to the 7.33 cm (mean of Figure 3.7) reported in this work. A larger joint position error can be the result of a smaller dataset and/or of training SINN/RNN for the upper/lower body separately. The mean joint position error for the proposed SINN approach is 6.23 cm. This error is smaller than both the reported error of the RNN approach in this work and that of Huang et al. However, these differences are small in magnitude, namely, approximately 1 cm. This indicates that the proposed SINN approach can provide an alternative for estimating full-body poses using only five IMUs for the RNN approach, requiring less computation power and training data.

Large jumps in outcome poses that were observed using a snapshot approach [210] have been reduced by using a stacked input vector, as can also be seen from the supplementary material. Furthermore, the mean joint position error is approxi-

mately 2 cm smaller than the one reported in [210]. In addition, the joint position error of individual joints showed a similar distribution to that observed in previous work. This was to be expected due to the kinematic chain that is evaluated, which effectively accumulates joint position errors from the proximal joints to the more distal joints. Joint jerk errors improved compared to previous work, which indicates that time coherency between outcome poses can be improved by stacking poses in the input vector. However, RNN results show smaller joint jerk errors than the SINN results, which shows that more time coherent outcomes can be obtained by explicit recurrent modeling.

The combined delay from the chosen configuration and reported computation time is 72 ms for the SINN approach and 117 ms for the RNN approach using a MATLAB implementation. These reported computation times are only indicative, since shorter delays are expected for a C++/firmware implementation. The SINN approach can be used to estimate full-body poses within acceptable delay boundaries (20–100 ms, according to [91]); however, the delay is close to the upper boundary. This can be improved by using less future information at the expense of accuracy. The RNN approach delay can be improved in a similar fashion or by using higher computational power. Alternatively, cloud computing could provide a more powerful computing environment without requiring such powerful equipment on site [121]. However, feasibility of such a solution for real-time pose estimation largely depends on the available internet speed.

Results presented in this work allow for decreasing the number of sensors/markers in full-body motion capture. While this approach is not tailored to any specific application, it was shown that cyclical and repetitive motions are estimated with the highest accuracy. Therefore, this approach has the largest potential to be applied to activities with these characteristics, e.g., biomechanical analysis of running [209], providing biofeedback to patients [99] or industry applications [66]. However, applications with less cyclical/repetitive motions may require more fine-tuning effort to reach the required level of accuracy.

3.4.1 Future Work

The results in this work were obtained by training a shallow/deep learning approach on a relatively small dataset. The effect of dataset size on performance of the proposed approaches remain unclear, and would require further analysis. The increased number of publicly available datasets[33, 122, 129] could provide opportunities for such an analysis.

The results in this work were based on inputs from the full-body motion capture output of Xsens MVN, while using orientations of five single IMUs directly might result in a decreased performance for estimating full-body poses. Therefore, additional research is required to evaluate the use of orientation and acceleration input of five single IMUs.

Performance of the proposed approaches varies with dynamics of the evaluated activities. A concept that could potentially improve this effect is to apply an adaptive time window based on the acceleration data, e.g., longer input sequences for low dynamic activities and the opposite for high dynamic activities. This concept can be applied to the RNN approach directly, while the SINN approach would require various trained networks (with different settings), which can then be chosen to use

at run-time depending on the dynamics of the activity.

The current analysis was performed using MATLAB; for development of a specific application, it should be preferred to use a different programming environment, such as Python or C++. This would likely also result in improved evaluation delays; however, the presented delay results can provide a benchmark.

3.5 Conclusions

The goal of this work was to evaluate the performance of estimating full-body poses using only five IMUs with a shallow (SINN) compared to a deep (RNN) learning approach. It has been shown that similar joint position accuracy (SINN $\approx$ 6 cm and RNN $\approx$ 7 cm) was achieved for both approaches with the considered dataset. However, the RNN approach results in smaller joint jerk errors, which is possibly the result of the explicit recurrency of the network. Furthermore, the SINN approach estimates poses with smaller delays, which allows for real-time applications. However, a SINN approach provides no flexibility to change the size and configuration of the stacked input vector at run-time. Therefore, choosing either approach would depend on several factors, namely, available computing power, dataset size, and/or real-time requirements.

3.6 Supplementary Material

A visualization of the results obtained from using either deep or shallow learning for estimating full-body poses using a minimal sensor set can be found at `https://youtu.be/XFXQWgFfpXo`:

Part II

Applications for Pose Estimation with Minimal Sensing

Chapter

FOUR

Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors

Published as:
F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. J. F. van Beijnum and P. H. Veltink **Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors** Frontiers in Physiology 9:218, pp. 1175-1180 http://doi.org/10.3389/fphys.2018.00218

Abstract

Analysis of running mechanics has traditionally been limited to a gait laboratory using either force plates or an instrumented treadmill in combination with a full-body optical motion capture system. With the introduction of inertial motion capture systems, it becomes possible to measure kinematics in any environment. However, kinetic information could not be provided with such technology. Furthermore, numerous body-worn sensors are required for a full-body motion analysis. The aim of this study is to examine the validity of a method to estimate sagittal knee joint angles and vertical ground reaction forces during running using an ambulatory minimal body-worn sensor setup. Two concatenated artificial neural networks were trained (using data from 8 healthy subjects) to estimate the kinematics and kinetics of the runners. The first artificial neural network maps the information (orientation and acceleration) of three inertial sensors (placed at the lower legs and pelvis) to lower-body joint angles. The estimated joint angles in combination with measured vertical accelerations are input to a second artificial neural network that estimates vertical ground reaction forces. To validate our approach, estimated joint angles were compared to both inertial and optical references, while kinetic output was compared to measured vertical ground reaction forces from an instrumented treadmill. Performance was evaluated using two scenarios: training and evaluating on a single subject and training on multiple subjects and evaluating on a different subject. The estimated kinematics and kinetics of most subjects show excellent agreement ($\rho > 0.99$) with the reference, for single subject training. Knee flexion/extension angles are estimated with a mean RMSE less than 5 degrees. Ground reaction forces are estimated with a mean RMSE less than 0.27 BW. Additionally, peak vertical ground reaction force, loading rate and maximal knee flexion during stance were compared, however, no significant differences were found. With multiple subject training the accuracy of estimating discrete and continuous outcomes decreases, however, good agreement ($\rho > 0.9$) is still achieved for seven of the eight different evaluated subjects. The performance of multiple subject learning depends on the diversity in the training dataset, as differences in accuracy were found for the different evaluated subjects.

Keywords: Machine Learning; Artificial Neural Networks; Reduced Sensor Set; Inertial Motion Capture; Running; Kinetics.

4.1 Introduction

Running is a very popular form of physical activity, that is often accompanied with a high occurrence of lower extremity injuries (incidence rate varies between 19.4 and 79.3 % [198]). It is assumed that there is a correlation between the development of these injuries and a runner's technique [76]. Additionally, improvements in running technique could lead to improved running performance [72, 112, 191]. Identifying the parameters in running technique that might be associated with injury development and/or running performance improvement requires a biomechanical analysis. This has traditionally been performed inside a gait laboratory using a three-dimensional optical motion capture system and force plates [136]. The most relevant kinematic and kinetic parameters analyzed are: joint angles [54, 63] and ground reaction forces [34], as these are important determinants of running technique [76]. Discrete kinetic parameters that are related to running injuries and/or performance are: loading rate and peak vertical ground reaction forces [171, 76, 46], whereas maximal knee flexion during stance is a relevant discrete kinematic parameter [63]. However, a lab setting is not identical to the regular running environment and may therefore result in different kinematics and kinetics [182]. Previous studies have confirmed this, showing significant differences between running on a treadmill and outdoors [135]. Furthermore, dissimilarities in running kinematics can also occur as a result of force plate targeting in overground lab running [37]. Therefore, a system capable of measuring relevant parameters outside of a laboratory may address these shortcomings.

Kinematic analysis can be performed in an ambulatory setting using inertial measurement units (IMUs) (see, for instance, [161]). [149] have used IMUs to measure the effects of fatigue on running mechanics during an actual marathon. However, this approach requires one sensor to be attached on each main body segment along a continuous 'kinematic chain', and therefore results in a large number of sensors and extensive subject preparation. Data driven approaches were shown to have potential for reducing the number of sensors in motion capture. [192] proposed a method for full-body motion capture by using a limited number of accelerometers; however, their nearest neighbor approach requires a database of prerecorded movements to be available at run-time. [210] showed comparable performance with a reduced sensor setup using an artificial neural network (ANN), trained to map five orientations to a full-body pose. ANNs have the advantage to create a 'model' for mapping certain inputs to outputs based on the dataset used for training [4]. Running applications using a minimal inertial sensor set have mainly focused on temporal outcomes, such as the use of gyroscopes on the feet to estimate temporal running parameters [125]. [15, 16] showed that with foot-mounted IMUs this can be extended to estimate spatio-temporal running parameters.

Ground reaction forces are also relevant outcome parameters for running analysis (e.g., [31, 34, 43, 136, 155]), since abnormal peak and/or loading rate values can lead to impact and overuse injuries, when the stress/frequency combination is above the runner's threshold [85, 127]. However, none of the aforementioned approaches provided users with kinetic information. Efforts to move kinetic analyses out of the laboratory setting have proven to be effective for trunk bending [65], gait [97], dance [176] and running [140]. However, aforementioned approaches require full-body kinematic information. The peak vertical ground reaction forces (vGRF) estimation approach of [39] relied only on tibial accelerations, but was not suitable

for estimation of kinetics during the whole stance phase. An approach relying only on trunk accelerations was not sufficient for vGRF estimation using a mass-spring-damper model [132].

To the best of our knowledge, there is no system that can provide runners with insights in both their kinematics and kinetics in an outdoor setting. The aim of this study is to assess the validity of a method to estimate knee joint angles and vertical ground reaction forces during running using an ambulatory minimal body-worn sensor setup. An ANN is trained to estimate joint angles based on lower leg orientations relative to the pelvis, similar to the approach presented in previous work [210]. Corresponding performance is evaluated using both inertial and optical full-body motion capture data. The estimated joint angles in combination with sensor accelerations can be fed into a second ANN which estimates (vertical) ground reaction forces. The proposed method was evaluated using continuous outcomes (vGRF and knee angle profiles) and discrete outcomes (peak vGRF, loading rate and maximal knee flexion during stance). The findings of this study could have potential for future applications in prevention of running injuries and improvement of running performance.

4.2 Materials and Methods

4.2.1 Measurement Protocol

Eight healthy experienced runners (8 males; age: 25.1 ± 5.2 years; height: 183.7 ± 4.5 cm; weight: 77.7 ± 9.4 kg; body mass index: 23.0 ± 2.5 kg/m^2) voluntarily participated in this research. The runners were recruited from a local track & field club and all reported no recent injuries. Subjects were instructed to run at 3 different speeds (10, 12 and 14 km/h, in this order) for 3 minutes each on an instrumented treadmill, located at the gait laboratory of the Roessingh Research and Development (Enschede, the Netherlands). A warm-up session at a self-selected running speed (of approximately 3 minutes) was performed by all subjects preceding the measurements. The ethics committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente approved this protocol and all subjects provided written informed consent prior to the measurements.

4.2.2 Measurement Setup

Reference kinematics were recorded with an optical motion capture system using the Plug-in Gait protocol[1] (Nexus 1.8.5, Vicon, Oxford, UK), with 41 retro-reflective markers placed directly on the runners' skin, as shown in Figure 4.1. The position of these markers was captured (at 100 Hz) by six high-speed infrared cameras (MX-13, Vicon, Oxford, UK) placed around the treadmill. Any object that could block the camera view or produce undesired reflections was removed from the measurement environment. Additionally, kinematics were synchronously captured using the Xsens MVN Link inertial motion capture system (Xsens, Enschede, the Netherlands), consisting of 17 IMUs placed at both shoulders, upper arms, lower arms, hands, upper legs, lower legs, feet, head, sternum, and pelvis [161]. The required

[1] `https://www.vicon.com/downloads/documentation/plug-in-gait-product-guide`

full-body Lycra suit (for IMU placement) was modified with holes to reduce motion artifacts of the retro-reflective markers, which are placed directly on the subject's skin. Full-body kinematics were exported using the accompanying software (MVN studio 4.3.7, Xsens, Enschede, the Netherlands) at a selected sampling frequency of 240 Hz. Subjects ran on a S-Mill instrumented treadmill (ForceLink, Culemborg, the Netherlands), with a running area of 250 x 100 cm, which can be seen in Figure 4.1. The treadmill was equipped with a 1-dimensional force plate, able to measure reference vGRF at 1000 Hz. Data of the different systems were synchronized using an analog synchronization signal.

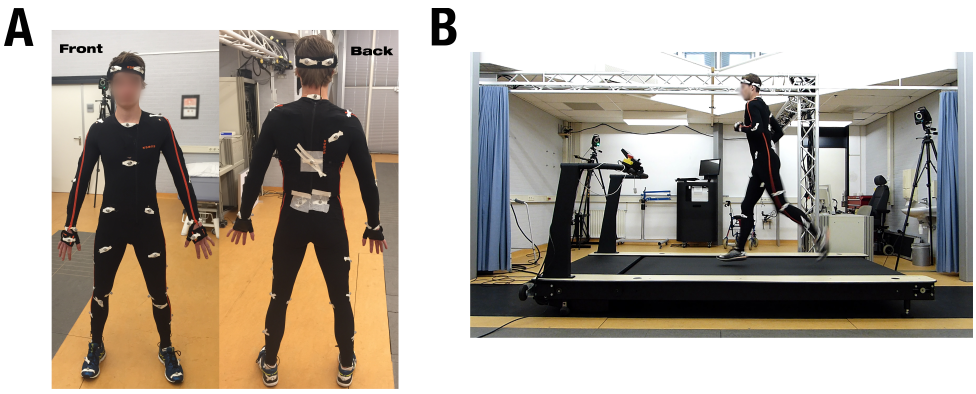


Figure 4.1: The measurement setup, (**A**) shows a front and back view of the sensor and retro-reflective marker placement (**B**) shows the measurement setup (only 2 cameras are visible in this angle). Subjects wore a Lycra suit to hold the IMUs in place, which was customized with holes to accommodate the placement of retro-reflective markers on the subject's skin. In this manner it was possible to measure kinematics simultaneously using both an inertial and optical motion capture system. The retro-reflective markers were placed according to the Plug-in Gait protocol. To ensure retro-reflective marker placement during the whole measurements, tapes were placed around these markers. *Note that written informed consent was provided for use of these images.*

4.2.3 Data Processing

The different trials were cropped to contain only kinematic and kinetic data of running at a steady speed, i.e., starting and stopping of the treadmill was disregarded. Optical kinematic data was processed using Plug-in Gait [49, 94]. The optical and inertial motion data did not require coordinate systems alignment as the outcome measures were expressed in the joint frame, according to ISB conventions [211]. The vGRFs were low-pass filtered at 20 Hz using a zero-phase 6th order Butterworth filter, to remove noise artifacts such as vibrations of the treadmill [184], while neither the optical nor inertial motion capture data were filtered. Beside the temporal alignment (achieved with an analog synchronization signal), the data were resampled at 120 Hz using linear interpolation (for the optical data) and downsampling (for the inertial and vGRF data), such that all synchronized data can be used in the proposed machine learning approach. This data resampling does not significantly influence the measured kinematics and kinetics, as was also concluded by [140]. For

analysis, the kinematic and kinetic data were segmented in stance phases using a 20 N threshold [128]. All data processing and statistical analyses was done in MATLAB R2017a (Mathworks, Inc., Natick, MA, USA).

4.2.4 Learning Approach

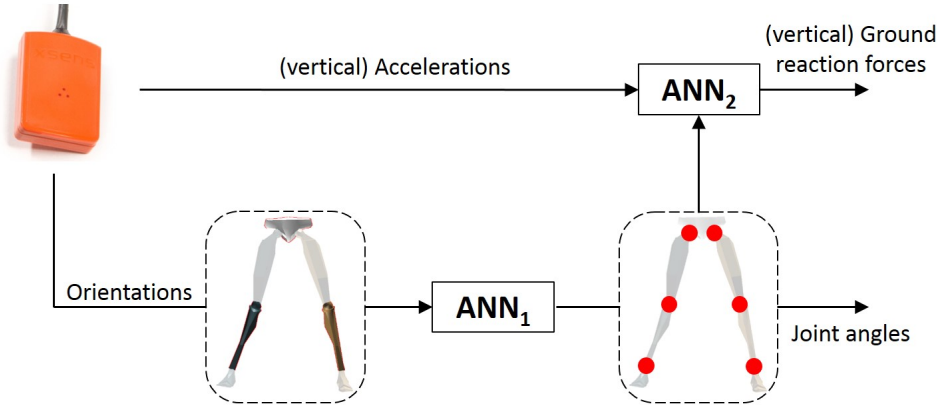


Figure 4.2: The IMU in the top left represents the sensors strapped to the lower legs and pelvis. Information from these sensors is used by two concatenated Artificial Neural Networks (ANNs) to estimate kinematics and kinetics. ANN_1 maps the relative orientations of the lower legs (with respect to the pelvis) to lower body joint angles (hip, knee and ankle). ANN_2 is trained to map the estimated kinematics in combination with the vertical (after transformation to the global frame) sensor accelerations to the reference ground reaction forces.

The proposed learning approach relies on data from 3 body-worn sensors (placed at the pelvis and lower legs), which are fed to a concatenation of two ANNs, as schematically represented in Figure 4.2. The first artificial neural network (ANN_1) maps relative (to the pelvis) orientations (in quaternions) of the lower legs to joints angles, whereas the second artificial neural network (ANN_2) maps the estimated joint angles in combination with vertical sensor accelerations (in the global frame) to vertical ground reaction forces. This architecture was chosen to allow for independent training of the two ANNs. Additionally, the proposed architecture separates the learning problems allowing for "selective" re-training of the ANNs (for instance, additional running environments can be included in the dataset of ANN_1 without measuring GRFs simultaneously).

Estimated kinematic outputs were being compared to measured reference kinematics, which were obtained from both inertial or optical motion capture systems. To that end, two training schemes were evaluated, as shown in Table 4.1, to test the proposed method irrespective of the motion capture technology.

Previous studies have achieved varying performance in GRF estimation [39, 65, 97, 132, 140, 176]. Therefore, several ANNs were trained using combinations of different input features (joint angles, pelvis and lower leg vertical accelerations) to select the best set of input features. The selection of these input features is based on their physical relation to the ground reaction forces, where joint angles define the

Table 4.1: The training and testing schemes for both the kinematic and kinetic estimations are represented. Where input to ANN_1 is in all cases the measured relative orientations from the three on-body IMUs (placed at the pelvis and both lower legs), and the output can be from the inertial (IMU) or optical (Plug-In Gait) measurements. This is then input to the kinetic estimation part (ANN_2), for which the output is all cases the measured vertical Ground Reaction Forces (vGRF) using the forceplates (FP).

ANN_1			
Scheme	**Training input**	**Training output**	**Reference**
1	3 IMU orientations	IMU lower-body joint angles	IMU lower-body joint angles
2	3 IMU orientations	Plug-In Gait lower-body joint angles	Plug-In Gait lower-body joint angles
ANN_2			
Scheme	**Training input**	**Training output**	**Reference**
1	IMU lower-body joint angles	vGRF (FP)	vGRF (FP)
2	Plug-In Gait lower-body joint angles	vGRF (FP)	vGRF (FP)

continuous kinematic chain [65, 97] and accelerations are related to force according to Newton's second law of motion.

In accordance with previous work of the authors [210], a two-layer (with 250 and 100 neurons) function fitting neural network architecture was used for both ANNs, capable of mapping non-linearities between input and output. The networks were trained for 2,000 iterations and training was stopped early if the gradient did not decrease for 6 consecutive iterations or if the gradient was smaller than 1×10^{-6}. The neural network toolbox of MATLAB R2017a (Mathworks, Inc., Natick, MA, USA) was used to design, train, and evaluate the ANNs described above.

Two different evaluation scenarios were evaluated to show single (Section 4.3.1) and multiple subject (Section 4.3.2) performance:

1. For each subject, evaluation was done using all running data at 12 km/h, while data with other speeds (i.e., 10 and 14 km/h) are used for training.

2. All data from one subject were used at turn for evaluation, while all data of remaining subjects were used for training. Note that, for the sake of simplicity, we will show only results corresponding to data of running at 12 km/h.

Scenario 1 would require every new user to perform a training phase. Scenario 2 could potentially produce a more generic model, although the lack of personalization of the network may result in decreased performance.

4.2.5 Outcome Measures

The performance of the proposed method was evaluated by comparing both discrete and continuous outcomes, as commonly done in similar works about biomechanical analysis of running [34, 46, 54, 63, 171]. For the knee flexion/extension (F/E) the similarity between the estimates and reference was calculated using the Pearson's correlation coefficient (ρ) and Root Mean Squared Error (RMSE) (as defined by [151]). The mean ρ over these different strides was calculated using a Fisher transformation to obtain a more representative average Pearson's correlation coefficient [45]. Additionally, the maximum knee F/E angle during the stance phase was evaluated using a paired t-test (significance level of 0.05) and Bland-Altman plot [24]. Estimated vGRFs (normalized to body weight (BW)) were also evaluated using both continuous (ρ and RMSE) and discrete metrics (loading rate and peak vGRF). The kinetic analysis was however limited to the stance phase of each leg (as there is no

contact during swing phase). Since the passive vGRF peak is not clearly defined for mid- or forefoot strikers, this event was determined using the peak acceleration from the lower leg IMUs [205]. Using this event the loading rate was calculated as the slope of vGRF between 20 and 80 percent of the passive vGRF peak time [46, 205].

4.3 Results

Section 4.3.1 shows performance of the proposed method for training and evaluating on a single subject, where the difference between both sets is the running speed (scenario 1). Section 4.3.2 is about generalization of this approach over different subjects (scenario 2).

4.3.1 Single subject learning

Kinematics estimation

The accuracy of estimated knee F/E angles based on different references (full-body IMU motion capture system or optical Plug-In Gait output) is presented in Table 4.2. The estimates provided by most individually trained ANNs have excellent agreement ($\rho > 0.99$) with the reference joint angles. Furthermore, only subject 8 shows significant differences in performance between the different references.

Table 4.2: Accuracy of estimated knee flexion/extension (F/E) angles (using ANN_1) with different training outputs (namely: IMU or Plug-in Gait-based), using single subject training and evaluation. Pearson's correlation coefficient (ρ) is calculated for each stride and averaged over approximately 200 strides for each subject (S01, S02, S03, S04, S05, S06, S07 and S08). The Root Mean Squared Error (RMSE) is calculated similarly over all strides. Training of the artificial neural networks was performed using running data at 10 and 14 km/h, while 12 km/h running data was used for evaluation.

Subjects	IMU				Plug-in Gait			
	Left F/E		Right F/E		Left F/E		Right F/E	
	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)
S01	0.99	3.24 (1.53)	0.99	4.38 (1.71)	0.99	3.56 (0.97)	0.99	4.76 (1.46)
S02	0.99	1.74 (0.48)	0.99	1.77 (0.54)	0.99	4.14 (1.39)	0.99	3.79 (1.41)
S03	0.99	2.65 (0.64)	0.99	2.05 (0.53)	0.99	3.70 (1.22)	0.99	2.58 (0.72)
S04	0.99	2.60 (0.47)	0.99	2.26 (0.58)	0.99	3.02 (1.28)	0.99	3.59 (1.41)
S05	0.99	3.39 (1.79)	0.99	3.55 (2.05)	0.99	4.03 (1.19)	0.99	4.49 (1.33)
S06	0.99	3.57 (0.67)	0.99	3.52 (0.64)	0.99	2.62 (0.54)	0.99	2.27 (0.63)
S07	0.99	3.30 (0.57)	0.99	2.86 (0.51)	0.99	5.27 (1.14)	0.99	5.41 (1.21)
S08	0.99	3.95 (1.70)	0.99	3.17 (1.49)	0.98	7.33 (2.68)	0.98	8.41 (3.02)

Mean (and standard deviation) of the estimated knee F/E angle profiles are shown in Figure 4.3 for a representative subject (S03). The largest difference between the estimate and its respective reference can be seen at the largest flexion angle, which is overestimated in all cases. As observed before in Table 4.2, differences between the estimates based on the various references are limited (4 degrees on average).

Table 4.3 shows the mean (and standard deviation) of the maximal knee F/E angle for each subject. Only inertial results and the corresponding estimates are presented in this table for conciseness. The mean difference in maximal knee flexion angle during stance between the estimate and its reference are less than 2 degrees for all subjects, and this result shows no significant differences ($p > 0.05$). A small bias of 0.4 degrees was found with limits of agreement -4.1 to 4.9 degrees for the

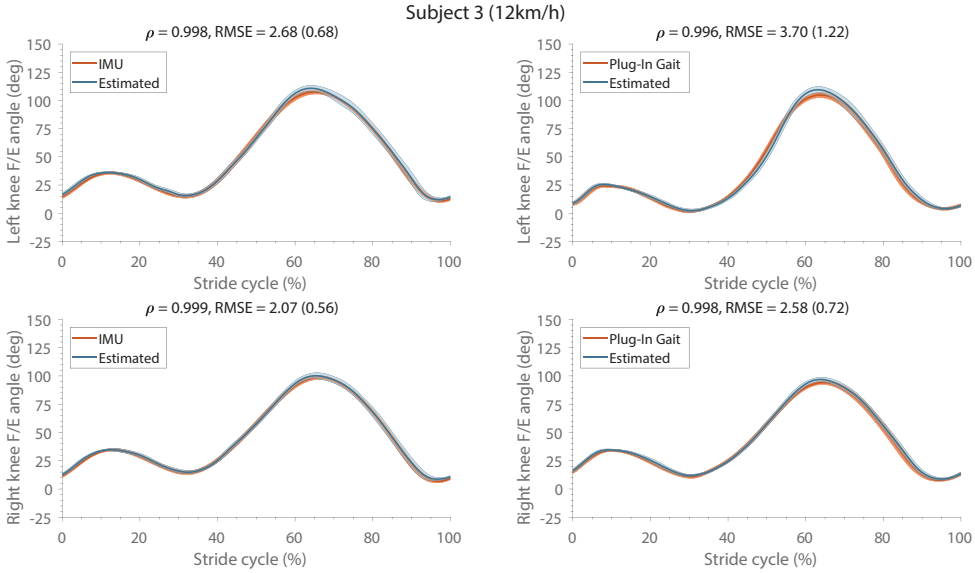


Figure 4.3: Mean (and standard deviation band) of the flexion/extension knee joint angle (in degrees) estimates are presented (normalized to the stride cycle) compared to their respective references (IMU and Plug-In Gait output). These estimates were obtained from training (using running data at 10 and 14 km/h) and evaluating (using running data at 12 km/h) on a single subject, similar results were obtained for the other subjects. The top row shows the angles of the left side and the bottom row presents the right side. At the top of each graph Pearson's correlation coefficient, root mean square error (RMSE) and the standard deviation (between the brackets) are specified, which were calculated for the estimate compared to its respective reference kinematics.

Table 4.3: The mean (and standard deviation) of discrete outcome measures for both the estimate and its corresponding reference (based on inertial full-body motion capture data) of all subjects. These estimates were obtained by training and evaluating on a single subject. Outcomes are averaged over approximately 400 steps (left and right combined). P-values are calculated using a paired t-test with the subject mean values.

Subjects	Max knee F/E angle (degrees)		vGRF peak (BW)		Loading rate (BW/s)	
	Reference	Estimate	Reference	Estimate	Reference	Estimate
S01	45.41 (3.56)	45.05 (3.94)	2.79 (0.08)	2.83 (0.06)	44.39 (7.37)	45.52 (8.05)
S02	42.96 (1.55)	42.56 (1.29)	2.96 (0.07)	2.94 (0.05)	50.72 (4.93)	46.38 (6.25)
S03	35.18 (1.25)	35.55 (1.06)	2.95 (0.08)	3.00 (0.08)	58.41 (6.86)	51.90 (7.47)
S04	41.11 (1.22)	41.68 (1.17)	2.81 (0.07)	2.82 (0.05)	56.97 (8.71)	50.55 (7.60)
S05	36.38 (2.08)	38.24 (4.36)	3.21 (0.10)	3.12 (0.08)	68.77 (7.65)	64.86 (7.44)
S06	35.12 (3.05)	34.30 (2.63)	3.01 (0.09)	3.07 (0.08)	48.56 (5.13)	53.81 (5.57)
S07	39.24 (1.92)	40.85 (2.37)	2.99 (0.08)	2.98 (0.12)	58.06 (8.37)	51.73 (6.10)
S08	39.45 (1.99)	39.58 (1.65)	3.02 (0.08)	3.02 (0.07)	47.92 (7.23)	44.88 (6.11)
Mean	39.36 (3.72)	39.72 (3.59)	2.97 (0.13)	2.97 (0.10)	54.23 (7.86)	51.20 (6.44)
p-value		0.31		0.79		0.08

comparison between the estimated maximal knee F/E angle during stance and the corresponding reference. Figure 4.4 (A) shows the related Bland-Altman plot. Occasional outliers (for three of the evaluated subjects) can be observed, which are mostly overestimating the maximal knee F/E angle during stance.

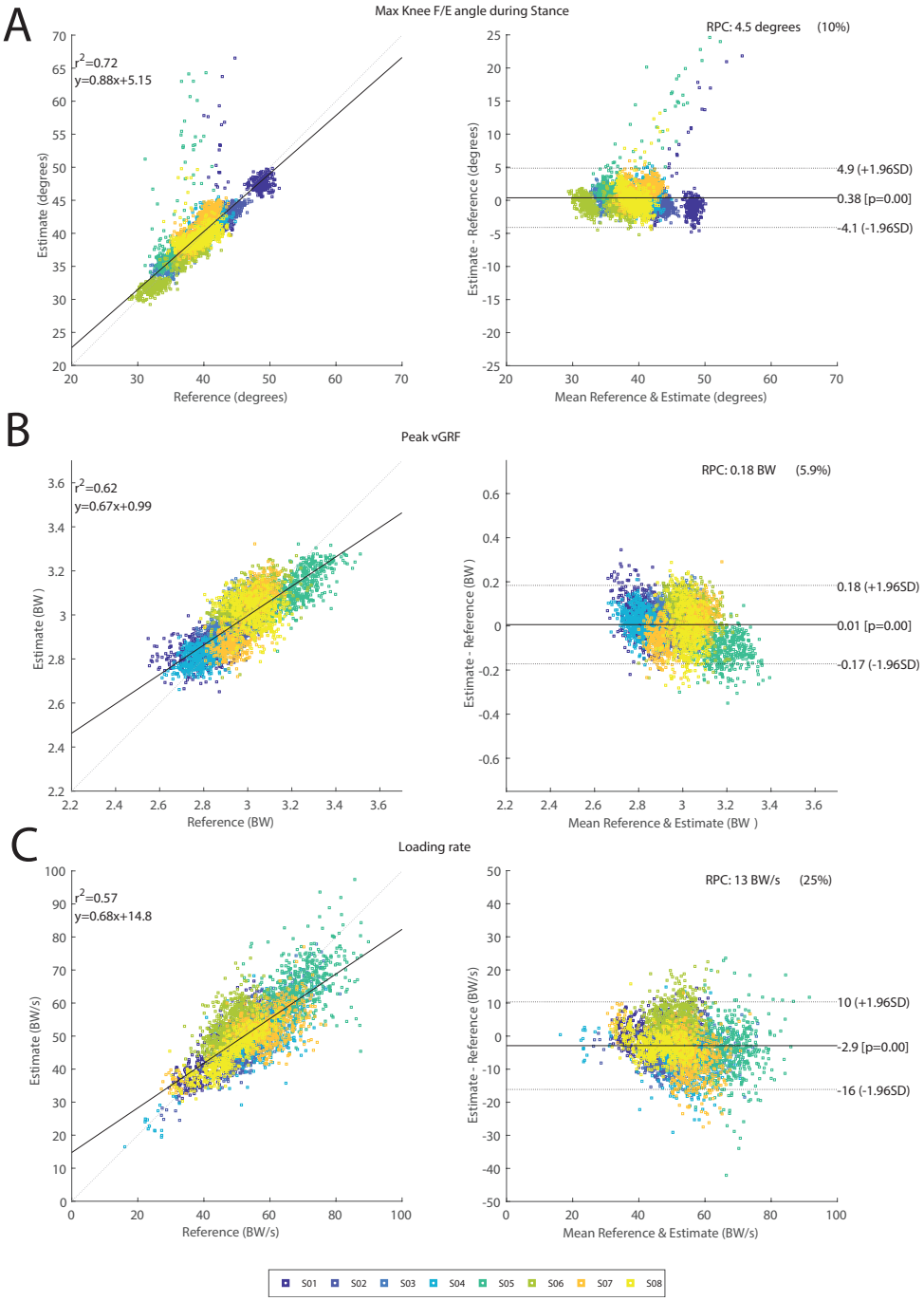


Figure 4.4: The left side shows the correlation plot of the discrete outcome measures: maximal knee flexion angle during stance (A), peak vGRF (B) and loading rate (C). The right side shows the corresponding difference plots of those three discrete outcome measures. Approximately 4,000 data points are shown, where different subjects are represented by the various colors.

Table 4.4: Accuracy of the estimated vertical ground reaction force (vGRF) using different input features (namely: joint angles ($\boldsymbol{\theta}_{joint}$), pelvis vertical acceleration ($\boldsymbol{a_P}$), all (pelvis, left and right lower leg) vertical accelerations ($\boldsymbol{a_{P+L}}$) or a combination of these). The evaluated set of features is shown above each column. These results were obtained using single subject training and evaluation. Pearson's correlation coefficient (ρ) is calculated for each contact and averaged over approximately 200 stance phases for each subject (S01, S02, S03, S04, S05, S06, S07 and S08). The Root Mean Squared Error (RMSE) is calculated similarly over all contacts, and the standard deviation of the RMSE is shown in brackets. The highest correlations (ρ) and smallest RMSE are shown in bold.

Subjects		**Features**									
		$\boldsymbol{a_P}$		$\boldsymbol{a_{P+L}}$		$\boldsymbol{\theta_{joint}}$		$\boldsymbol{a_P}$ & $\boldsymbol{\theta_{joint}}$		$\boldsymbol{a_{P+L}}$ & $\boldsymbol{\theta_{joint}}$	
		ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)
S01	L vGRF	0.97	0.26 (0.03)	0.98	0.21 (0.03)	0.99	0.20 (0.06)	0.99	0.15 (0.04)	**0.99**	**0.12 (0.03)**
	R vGRF	0.96	0.26 (0.02)	0.99	0.17 (0.02)	0.99	0.21 (0.05)	0.99	0.16 (0.04)	**0.99**	**0.15 (0.04)**
S02	L vGRF	0.94	0.33 (0.03)	0.97	**0.23 (0.02)**	**0.98**	0.24 (0.07)	0.97	0.24 (0.05)	0.97	0.25 (0.03)
	R vGRF	0.97	0.27 (0.03)	**0.97**	**0.25 (0.03)**	0.96	0.29 (0.08)	0.96	0.28 (0.06)	0.97	0.25 (0.04)
S03	L vGRF	0.96	0.29 (0.03)	0.99	0.14 (0.03)	0.99	0.20 (0.07)	0.99	0.12 (0.04)	**0.99**	**0.10 (0.03)**
	R vGRF	0.95	0.32 (0.03)	0.99	0.11 (0.03)	0.99	0.16 (0.05)	0.99	0.10 (0.03)	**0.99**	**0.09 (0.02)**
S04	L vGRF	0.96	0.25 (0.03)	**0.99**	**0.14 (0.03)**	0.99	0.16 (0.05)	0.99	0.17 (0.04)	0.99	0.15 (0.04)
	R vGRF	0.96	0.27 (0.03)	0.99	0.13 (0.04)	0.99	0.20 (0.07)	0.99	0.13 (0.04)	**0.99**	**0.11 (0.03)**
S05	L vGRF	0.97	0.28 (0.04)	**0.98**	**0.25 (0.05)**	0.95	0.37 (0.11)	0.97	0.30 (0.07)	0.97	0.30 (0.07)
	R vGRF	0.98	0.27 (0.03)	**0.98**	**0.25 (0.05)**	0.93	0.44 (0.14)	0.96	0.33 (0.08)	0.96	0.33 (0.08)
S06	L vGRF	0.95	0.32 (0.03)	**0.98**	**0.22 (0.04)**	0.95	0.38 (0.09)	0.96	0.34 (0.07)	0.96	0.30 (0.05)
	R vGRF	0.94	0.35 (0.04)	**0.98**	**0.21 (0.04)**	0.93	0.42 (0.10)	0.95	0.36 (0.06)	0.95	0.33 (0.05)
S07	L vGRF	0.91	0.40 (0.30)	**0.96**	**0.27 (0.04)**	0.93	0.38 (0.10)	0.96	0.30 (0.06)	0.96	0.28 (0.05)
	R vGRF	0.93	0.38 (0.03)	0.96	0.28 (0.04)	0.96	0.33 (0.08)	0.96	0.29 (0.07)	**0.97**	**0.25 (0.06)**
S08	L vGRF	0.97	0.27 (0.02)	0.99	0.12 (0.02)	0.99	0.20 (0.07)	0.99	0.14 (0.04)	**0.99**	**0.11 (0.03)**
	R vGRF	0.96	0.28 (0.03)	0.99	0.14 (0.03)	0.98	0.24 (0.07)	0.99	0.19 (0.05)	**0.99**	**0.12 (0.04)**

Kinetics estimation

Table 4.4 shows an overview of performance when different combinations of input features (joint angles, pelvis and lower leg accelerations) are evaluated. On average the best results (marked in bold for individual subjects) were achieved using a combination of all vertical accelerations and joint angles as input features. Therefore, results presented below are obtained when ANN_2 was trained using these features.

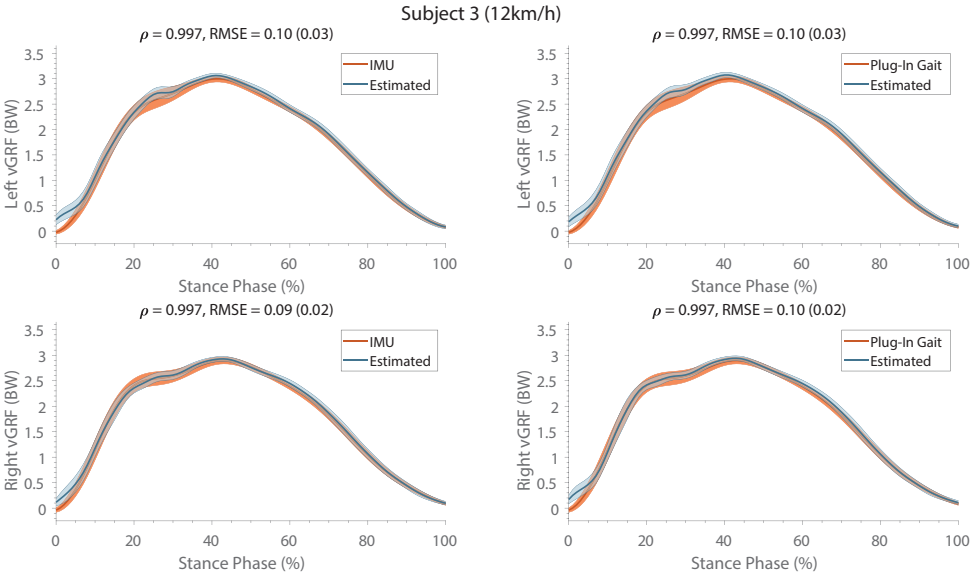


Figure 4.5: Mean (and standard deviation band) of the estimated ground reaction forces (in BW) are presented (normalized to the stance phase) compared to their respective references (IMU and Plug-In Gait joint angle output). These estimates were obtained from training and evaluating on a single subject, similar results were obtained for the other subjects. The top row shows the forces of the left contacts and the bottom row presents the right contacts. At the top of each graph Pearson's correlation coefficient, root mean square error (RMSE) and the standard deviation (between the brackets) are specified, which were calculated for the estimate compared to its respective reference kinematics.

The estimated ground reaction profiles of a representative subject (S03) are shown in Figure 4.5 for ANN_2 based on both reference kinematics (IMUs and Plug-In Gait). Similarly to what was observed for the estimated knee F/E angles, differences between the networks (ANN_2) trained on the various references are minimal. Largest differences between the estimated and reference vGRF can be seen at the beginning of stance phase. However, peak values are estimated with high accuracy, resulting in correlation coefficients larger than 0.96.

Results for the discrete outcomes (peak vGRF and loading rate) can be found in Table 4.3. Mean peak vGRF differences between the estimate and its reference are within 0.09 BW for all subjects, which resulted in no significant differences ($p > 0.05$). Variation between the estimate and its reference is larger for the loading rate, however this difference is still not significant ($p > 0.05$). Figure 4.4 (B and C) show the Bland-Altman plots for both the peak vGRF and loading rate. A small bias

of 0.01 BW is present in the estimated peak vGRF, with limits of agreement -0.17 to 0.18 BW. The loading rate is estimated with a bias of -2.9 BW/s with limits of agreement -16 to 10 BW/s. Both plots show occasional outliers for multiple subjects.

Variation in running speeds

Extrapolation capabilities of the proposed approach were investigated by evaluating different running speeds for subject 3. Figure 4.6 shows RMSEs for the evaluated speeds, where the remaining trials are in the training dataset. This figure shows that the most accurate continuous estimation can be achieve when an intermediate speed (12 km/h) is used, rather than the ones which are slower (10 km/h) or faster (14 km/h) than those in their respective training datasets.

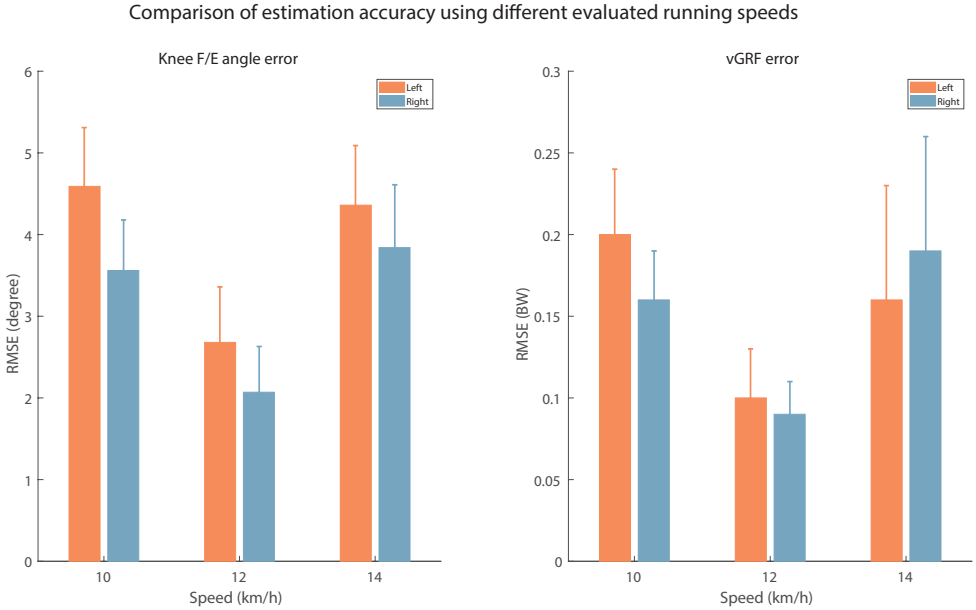


Figure 4.6: Accuracy of the estimated vertical ground reaction force (vGRF) and knee flexion/extension (F/E) angle for different evaluated speeds, hence the other speeds are part of the training dataset, using single subject training and evaluation, as described in Section 4.2.4. The artificial neural networks were trained with and evaluated relative to a full-body inertial kinematic measurement (Table 4.1, training scheme 1). The results for a representative subject are shown in this graph. The Root Mean Squared Error (RMSE) is calculated over all stride/stance phases and averaged over approximately 200 strides for each different evaluated speed (10, 12 and 14 km/h).

Additionally, discrete outcome measures were evaluated for the same subject, which are presented in Table 4.5. The peak vGRF and maximal knee flexion during stance also show that interpolating speeds results in more accurate outcomes than extrapolating. However, this trend is not present for the loading rate accuracy.

Table 4.5: The variation in discrete outcome measures for different speeds in subject 3. The mean (and standard deviation) of peak vGRF, loading rate and max knee flexion during stance are shown for both the estimate and its corresponding reference (based on inertial full-body motion capture data), these are calculated over approximately 400 steps (left and right combined). The artificial neural networks were trained using running data of two speeds (different from the evaluation speed), while the shown speed was used for evaluation.

Parameter	Speed	Reference (IMU)		Estimate	
		Left	Right	Left	Right
Max knee flexion (degrees)	10 km/h	34.49 (1.10)	33.90 (1.20)	30.57 (0.98)	30.24 (1.22)
	12 km/h	35.64 (1.20)	34.71 (1.13)	36.18 (0.71)	34.92 (0.98)
	14 km/h	36.93 (1.24)	35.11 (1.16)	38.22 (1.58)	36.23 (2.76)
peak vGRF (BW)	10 km/h	2.85 (0.06)	2.77 (0.06)	2.76 (0.07)	2.67 (0.08)
	12 km/h	3.00 (0.06)	2.90 (0.07)	3.06 (0.04)	2.93 (0.05)
	14 km/h	3.13 (0.07)	3.00 (0.07)	2.96 (0.13)	2.92 (0.10)
Loading rate (BW/s)	10 km/h	52.92 (5.82)	55.11 (6.10)	46.05 (5.92)	56.05 (8.37)
	12 km/h	55.47 (6.12)	61.34 (6.29)	47.67 (6.28)	56.13 (6.03)
	14 km/h	63.25 (7.31)	67.17 (9.78)	59.11 (9.06)	55.52 (13.91)

Table 4.6: Accuracy of the estimated knee flexion/extension (F/E) angles (by ANN_1) and vertical ground reaction forces (vGRF) (by ANN_2) using different training outputs (namely: IMU or Plug-in Gait-based) by training on data of all subjects except for one which is used for the evaluation at 12 km/h. Pearson's correlation coefficient (ρ) is calculated for each stride and averaged over approximately 200 strides for each different test subject (S01, S02, S03, S04, S05, S06, S07 and S08). The Root Mean Squared Error (RMSE) is calculated similarly over all strides.

Knee F/E angle accuracy								
Subjects	IMU				Plug-in Gait			
	Left F/E		Right F/E		Left F/E		Right F/E	
	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)
S01	0.88	19.11 (4.92)	0.83	19.47 (3.66)	0.77	25.05 (2.20)	0.83	23.57 (2.13)
S02	0.99	8.57 (0.74)	0.99	8.09 (0.79)	0.98	11.87 (1.08)	0.98	6.76 (0.71)
S03	0.95	14.92 (1.54)	0.94	11.08 (1.32)	0.91	15.19 (1.54)	0.91	22.57 (4.03)
S04	0.98	8.35 (0.76)	0.98	6.68 (1.12)	0.93	11.36 (0.93)	0.98	6.90 (0.94)
S05	0.98	9.89 (1.10)	0.98	7.00 (1.28)	0.96	19.62 (3.62)	0.97	7.41 (1.43)
S06	0.98	7.33 (1.00)	0.99	6.99 (1.07)	0.97	7.70 (0.99)	0.98	8.76 (1.46)
S07	0.98	5.88 (0.68)	0.99	4.83 (0.99)	0.98	6.83 (0.81)	0.98	6.62 (0.85)
S08	0.98	6.36 (1.18)	0.99	4.66 (0.92)	0.98	6.29 (1.09)	0.99	7.21 (0.89)
vGRF accuracy								
Subjects	IMU				Plug-in Gait			
	Left vGRF		Right vGRF		Left GRF		Right vGRF	
	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)	ρ	RMSE (σ)
S01	0.92	0.45 (0.10)	0.90	1.25 (0.25)	0.94	0.52 (0.11)	0.86	0.56 (0.08)
S02	0.95	0.31 (0.08)	0.99	0.16 (0.03)	0.98	0.22 (0.04)	0.97	0.27 (0.05)
S03	0.95	0.50 (0.18)	0.98	0.60 (0.09)	0.83	0.81 (0.19)	0.97	0.38 (0.07)
S04	0.98	0.26 (0.06)	0.95	0.32 (0.05)	0.95	0.34 (0.19)	0.98	0.30 (0.05)
S05	0.99	0.21 (0.05)	0.97	0.32 (0.07)	0.97	0.33 (0.08)	0.99	0.20 (0.07)
S06	0.98	0.25 (0.04)	0.94	0.36 (0.03)	0.97	0.28 (0.04)	0.99	0.20 (0.04)
S07	0.96	0.30 (0.04)	0.98	0.22 (0.04)	0.97	0.29 (0.05)	0.97	0.27 (0.05)
S08	0.93	0.46 (0.05)	0.98	0.24 (0.04)	0.91	0.44 (0.05)	0.98	0.28 (0.04)

4.3.2 Multiple subject learning

The generalization performance of both ANNs were evaluated by training with all different combinations of subjects in the training and evaluation datasets. Table 4.6 (top-half) shows the results of kinematics for the different evaluated subjects. Seven out of the eight subject show correlations larger than 0.9, indicating good agreement. However, the RMSE is expectantly larger than for single subject learning (Section 4.3.1). The estimated knee F/E angles for subjects 1 and 3 are significantly less accurate. Additionally, the mean estimated knee F/E angle profiles of subject 4 are shown in Figure 4.7, with the measured references used for comparison. The stance phase (until approximately 30% of the stride cycle) is estimated with higher

accuracy than the swing phase, same behavior can be seen for single subject learning (Figure 4.3).

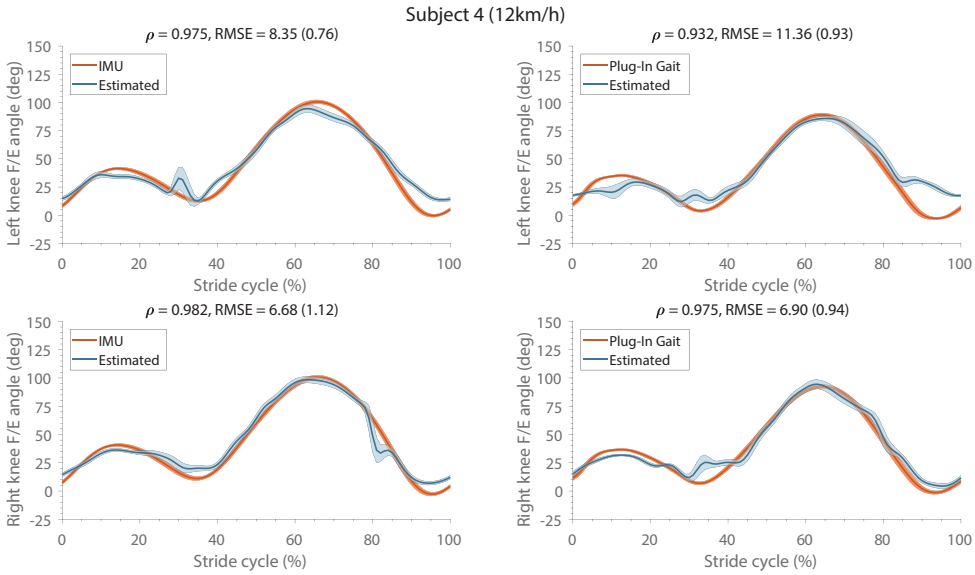


Figure 4.7: Mean (and standard deviation band) of the flexion/extension knee joint angle (in degrees) estimates are presented (normalized to the stride cycle) compared to their respective references (IMU and Plug-In Gait joint angle output). These estimates were obtained from training on multiple subjects and evaluating on a different subject, and were comparable to the other evaluated subjects. The top row shows the angles of the left side and the bottom row presents the right side. At the top of each graph Pearson's correlation coefficient, root mean square error (RMSE) and the standard deviation (between the brackets) are specified, which were calculated for the estimate and its respective reference kinematics.

Results of the kinetic estimations can be seen in Table 4.6 (bottom-half). Similar to the joint angles, vGRFs are mostly estimated with correlations larger than 0.9 indicating good agreement with the measurements. However, subjects 1 and 3 show lower correlation coefficients, as was also seen for the kinematics. Vertical ground reaction force profiles of one representative subject (S04) are shown in Figure 4.8, which shows an increase in RMSEs compared to the single subject learning (Figure 4.5). The maximum estimated ground reaction forces are mostly comparable to the reference.

The accuracy of estimating discrete outcome measures is shown in Table 4.7. The estimation accuracy varies between different subjects and outcome measures. However, in most cases an increase in error can be seen when comparing to the single subject training (Table 4.4). Additionally, an increase in the standard deviations of the different estimated outcome measures can be seen. However, the estimated outcome measures and the corresponding references were not found to be significantly different.

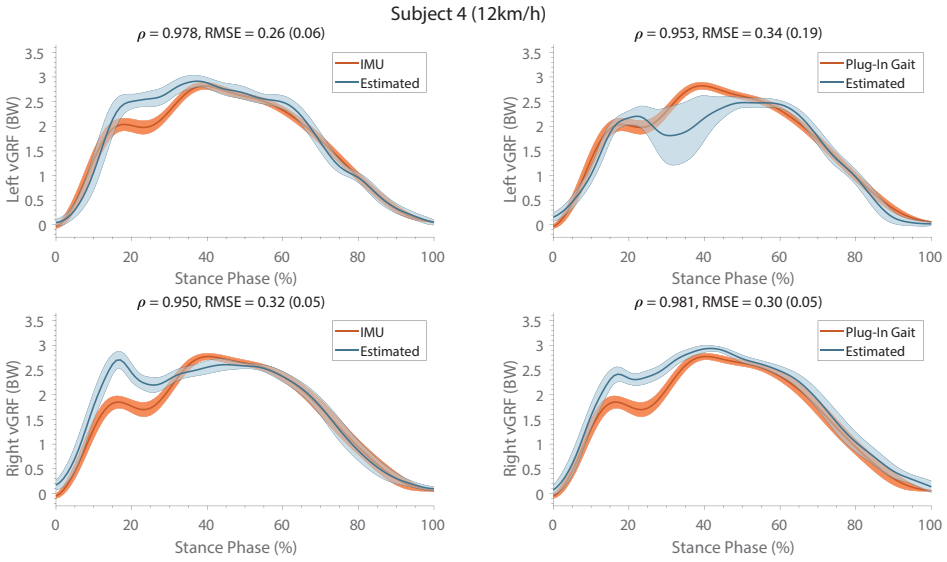


Figure 4.8: Mean (and standard deviation band) of the estimated vertical ground reaction forces (in BW) are presented (normalized to the stance phase) compared to the measured reference. These estimates were obtained from training on multiple subjects and evaluating on a different subject, and were comparable to the other evaluated subjects. The top row shows the forces of the left contacts and the bottom row presents the right contacts. At the top of each graph Pearson's correlation coefficient, root mean square error (RMSE) and the standard deviation (between the brackets) are specified, which were calculated for the estimate and its respective reference kinematics.

Table 4.7: The mean (and standard deviation) of discrete outcome measures for both the estimate and its corresponding reference (based on inertial full-body motion capture data) of all subjects. These estimates were obtained by training on multiple subjects and evaluating on a different subject (using running data at 12 km/h). Outcomes are averaged over approximately 400 steps (left and right combined). P-values are calculated using a paired t-test with the subject mean values.

Subjects	Max knee F/E angle (degrees)		vGRF peak (BW)		Loading rate (BW/s)	
	Reference	Estimate	Reference	Estimate	Reference	Estimate
S01	45.41 (3.56)	65.90 (24.61)	2.79 (0.08)	3.82 (0.96)	44.39 (7.37)	33.86 (35.59)
S02	42.96 (1.55)	33.04 (1.19)	2.96 (0.07)	2.92 (0.13)	50.72 (4.93)	43.91 (7.38)
S03	35.18 (1.25)	40.21 (5.38)	2.95 (0.08)	2.72 (0.64)	58.41 (6.86)	38.72 (8.27)
S04	41.11 (1.22)	38.35 (3.73)	2.81 (0.07)	2.86 (0.14)	56.97 (8.71)	70.75 (12.11)
S05	36.38 (2.08)	41.88 (3.81)	3.21 (0.10)	3.15 (0.19)	68.77 (7.65)	48.87 (17.20)
S06	35.12 (3.05)	37.42 (2.68)	3.01 (0.09)	3.01 (0.12)	48.56 (5.13)	41.47 (3.58)
S07	39.24 (1.92)	38.97 (3.65)	2.99 (0.08)	2.91 (0.08)	58.06 (8.37)	58.19 (8.75)
S08	39.45 (1.99)	38.36 (1.68)	3.02 (0.08)	3.17 (0.23)	47.92 (7.23)	56.06 (10.79)
Mean	39.36 (3.72)	41.77 (10.08)	2.97 (0.13)	3.07 (0.34)	54.23 (7.86)	48.98 (12.10)
p-value		0.47		0.26		0.37

4.4 Discussion

This work shows that sagittal knee kinematics and vGRF can be estimated using only three inertial sensors placed on the lower legs and pelvis, in particular, the peak vGRF, maximal knee F/E angles during stance, and the knee F/E angles and vGRF profiles are estimated with no significant differences with respect to the reference.

Estimation of joint angles for a single subject has shown to be more accurate

(average RMSE < 5 degrees) than was achieved in previous work of the authors (average RMSE ≈ 7 degrees) [210]. This can partly be explained by the difference in composition of the training databases between both methods, since the current dataset had less variation of motions, i.e., only running. This approach requires obtaining reference kinetics and kinematics of each subject, i.e., each subject has to run on an instrumented treadmill.

Additionally, multiple subject learning results showed good agreement ($\rho > 0.9$) for most subjects in the continuous outcomes. However, the ANNs could not generalize over all idiosyncrasies of the individual subjects as RMSEs and differences in discrete outcomes increased, expectantly. Subjects had different landing patterns (heel, mid- or forefoot striking), which may be a reason for the degraded performance shown for example in subject 1. By including more subjects different models could be trained for each different landing phenotype. Alternatively, larger soft-tissue artifacts of the inertial sensors compared to the other subjects may explain the degraded performance.

No significant differences were found between any of the reference and estimated discrete outcome measures, for both evaluation scenarios. However, the required accuracy would largely be defined by the application of interest. An example of such an application could be tracking kinematic/kinetic changes due to fatigue, since they may relate to increased chance of injury [149]. However, more data (specific for such an application, e.g., running under fatigue) should be acquired to evaluate if the proposed approach can track such differences.

The running mechanics in this work are estimated based on inertial or optical motion capture data. Each of these technologies have their advantages and disadvantages [70]. Differences in the reference knee F/E profiles for the different technologies are observed for the results in Section 4.3.1, which can be explained by differences in the underlying models of the human body and their assumptions [95]. However, the estimated kinematics based on the different technologies are similar to their respective measured kinematics. This shows that the method has potential to be applied in this context irrespective of the preferred technology for recording training data. Therefore, the proposed method has potential to estimate output based on other kinematic references, such as biomechanical models driven by optical data [51, 188].

The measured dataset contains only treadmill running, however, the proposed method is not limited to be applied under these conditions. Evaluating the proposed method in a different setting (e.g., outdoor running) might result in less accurate estimations of knee F/E angles and vGRFs. To improve such results, the dataset can be extended by including running at different slopes of the treadmill. Furthermore, 3D ground reaction forces could be measured using pressure insoles for example [162], which enables data collection in any running environment for training data collection. Extrapolating kinematic and kinetic data outside of the training dataset appears to be more difficult than interpolating such data. This was shown by the degraded performance after training with different running speeds or extrapolating over various subjects. This indicates that careful construction of the training dataset is required to obtain the best possible performance.

A limitation of the proposed method is that only vertical kinetics can be estimated. This can be contributed to the available measurement setup, since it would require a treadmill instrumented with a force plate that can measure three-dimensional forces. However, our proposed method could be extended using the

three-dimensional GRF estimation approach of [97] using full-body inertial motion capture. Furthermore, only sagittal plane knee kinematics could be estimated in the proposed approach, possibilities of estimating kinematics of other joints and/or planes would require additional research.

The concatenated ANN approach allows for training the ANN_1 (kinematics) independent of the ANN_2 (kinetics). This enables the use of only inertial motion capture data in various environments for training ANN_1. Instead of concatenating two ANNs, a single ANN could be trained to map relative orientations and vertical accelerations to ground reaction forces and joint angles. Initial tests show comparable results for single subject training, however, multiple subject training was less successful. When one ANN is trained to estimate both kinematics and kinetics, cross-dependencies between features and outputs become important, which is less so for concatenated ANNs. This can be seen in the differences in accuracy between estimation of kinematics (ANN_1) and kinetics (ANN_2) for multiple subject training in Section 4.3.2.

Figure 4.5 shows differences in the measured reference vGRF between left and right stance phases, which can also be seen from the estimated output. This could indicate that the proposed method is capable of detecting differences between left and right kinetics. Note that, given the relatively short duration of the running sessions, effects of fatigue could not be evaluated using the current setup, but it is an interesting future development.

The estimated vertical ground reaction forces ($\rho > 0.99$ and RMSE < 0.27 BW) using the proposed method are comparable to that of [65] ($R^2 > 0.981$ and RSME < 10 N), who estimated GRFs during a bending task by using a full-body inertial motion capture system. [97] evaluated a similar approach on walking using inertial sensors, where the errors are comparable to the ones reported in the proposed method. [39] showed that by exploiting only tibial accelerations to estimate peak vGRFs an approximate RMSE of 6% can be achieved, however this method was only applied to training and testing on individual subjects. [176] estimated vGRF more accurately (3% error) than the proposed method, by relying on full-body optical motion capture for their method. [140] reported similar performance in estimation of the loading rate, while our proposed method was shown to estimate peak vGRFs more accurately. [39] reported peak vGRF estimation errors of approximately 6%, whereas our proposed method is able to estimate peak vGRF with an accuracy of less than 0.10 BW ($\approx 3.5\%$).

4.5 Conclusions

This work has shown the potential of estimating kinetics (vGRF) and kinematics (knee F/E angles) during running using a minimal on-body sensor setup (namely, 3 sensor devices placed on the lower legs and pelvis). Best performance can be obtained when the proposed approach is applied to a single subject. Training over multiple subjects was shown to be possible, since good agreement between the estimates and references were achieved, however the RMSEs are larger than for single subject training. In other words, the proposed method has potential to be applied for individual subjects, and with additional research can be extended for running in various environments.

4.6 Acknowledgments

The authors would like to thank the Roessingh Research & Development for the availability of the gait laboratory for the measurements. In particular the lab manager, Leendert Schaake, who helped significantly with the measurement setup and optical data processing.

Chapter

FIVE

Foot Progression Angle Estimation Using A Single Foot-Worn Inertial Sensor

Submitted as:
F. J. Wouda, S.L.J.O. Jaspar, J. Harlaar, B. J. F. van Beijnum and P. H. Veltink **Foot Progression Angle Estimation Using A Single Foot-Worn Inertial Sensor** IEEE Transactions on Biomedical Engineering

Abstract

Objective: The foot progression angle is an important measure used to help patients reduce their knee adduction moment. Current measurement systems are either lab-bounded or do not function in all environments (e.g., magnetically distorted). This work proposes a novel approach to estimate foot progression angle using a single foot-worn inertial sensor (accelerometer and gyroscope). *Methods:* The approach uses a dynamic step frame that is recalculated for the stance phase of each step to calculate the foot trajectory relative to that frame, to minimize effects of drift and to eliminate the need for a magnetometer. The FPA is then calculated as the angle between walking direction and the dynamic step frame. This approach was validated by gait measurements with five subjects walking with three gait types (normal, toe-in and toe-out). *Results:* The FPA was estimated with a maximum mean error of $\sim 2.6^\circ$ over all gait conditions. Additionally, the proposed inertial approach can significantly differentiate between the three different gait types. *Conclusion:* The proposed approach can effectively estimate differences in FPA without requiring a heading reference (magnetometer). *Significance:* This work enables feedback applications on FPA for patients with gait disorders that function in any environment, i.e. outside of a gait lab or in magnetically distorted environments.

Keywords: Foot Progression Angle, Inertial Sensors, Knee Osteoarthritis, Minimal Sensing, Zero Velocity Update, PCA

5.1 Introduction

Knee osteoarthritis (KOA) is among the most reported musculoskeletal diseases (men 10.1%, women 13.6%) and the leading cause for disability among the elderly [141, 207]. This disease has no cure currently, however, patients can make use of surgical, pharmacological and biomechanical treatments to improve their quality of life [156]. Pharmalogical treatment can reduce the effects of symptoms of KOA, in severe stages of the disease surgical treatment (knee replacement) could be considered [99]. Biomechanical treatment can help to reduce the knee loading, which has been shown to correlate with pain, cartilage degeneration and disease progression [150].

Biomechanical treatment can be achieved by use of braces, canes and/or gait retraining. No additional devices are required for gait retraining, however as this treatment is time consuming and space-bounded it has not been adopted on a large scale [19]. The goal of gait retraining is to reduce the loading on the knee by gradually modifying the patients' gait pattern [19]. Directly measuring the medial knee loading would require invasive force sensors and is therefore only possible after a knee replacement [58]. Alternatively, the medial knee loading can be estimated using a surrogate measure, namely the Knee Adduction Moment (KAM) [111]. The KAM can be estimated using inverse dynamics, which requires a full-body motion capture system and force measurements [177].

However, the KAM is not an optimal parameter to provide feedback to patients, since the relation to kinematic parameters is not evident to them [154]. Therefore, instructing patients using a kinematic adaptation (toe-in gait) results in more effective decrease of the KAM [179, 163, 178]. This can be quantified using the Foot Progression Angle (FPA), which is defined as the angle between the heading direction and foot orientation.

A recent study has shown that the FPA can effectively be measured using one foot-worn sensor [87], consisting of an accelerometer, gyroscope and magnetometer. However, the use of a magnetometer limits applications of this approach, since it requires a minimally disturbed Earth magnetic field. In various environments this is not the case, due to ferro-magnetic materials present in floors and walls [50]. With inaccurate measurements of the Earth magnetic field, no proper reference frame can be determined, hence inaccurate estimates of the FPA are obtained.

To the best of our knowledge, there is no single-sensor approach for estimating FPA in any environment (including magnetically disturbances). This resulted in the following aim of this study: design and evaluation of an approach to estimate FPA using a single foot-worn inertial sensor (accelerometer and gyroscope). Due to using a dynamic foot reference frame instead of an Earth reference frame no magnetometer is required. To minimize the effects of drift during a single step, the Zero Velocity Update (ZUPT) is applied [164]. The accuracy of the proposed method is validated using an optical reference system. The findings of this study could have potential for future applications in feedback systems for KOA patients.

5.2 FPA estimation

Our proposed FPA estimation approach consists of five steps as schematically displayed in Figure 5.1. The approach uses a dynamic foot frame (as schematically displayed in Figure 5.2), which changes from stance phase i to next stance phase $i+1$. This is done by integration of angular velocity during a step in between subsequent stance phases [28], updated with Zero Angular Velocity Update (ZAVU). Therefore, the start and end of a step should be determined using a zero-velocity detection [183, 92]. With strap-down integration, the vector from calcaneus position in stance phase i to calcaneus position in stance phase $i+1$ is determined in this dynamic foot frame. Subsequently correcting for drift using ZUPT and zero vertical position at the start and end of the step. FPA is calculated from the angle between foot direction during stance phase i and direction of the next step [163], which is estimated based on the endpoint of the trajectory estimation, as schematically displayed in Figure 5.2.

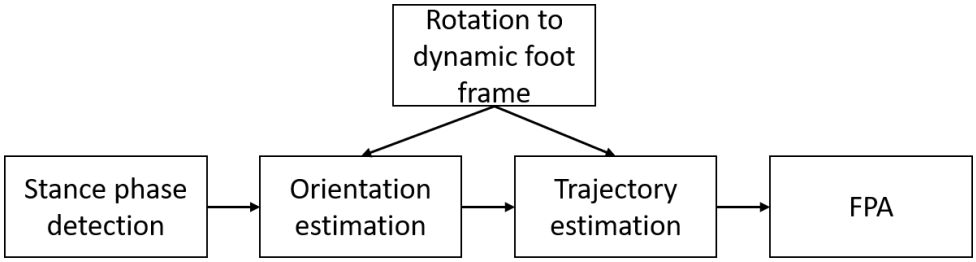


Figure 5.1: Flowchart of the proposed FPA estimation algorithm.

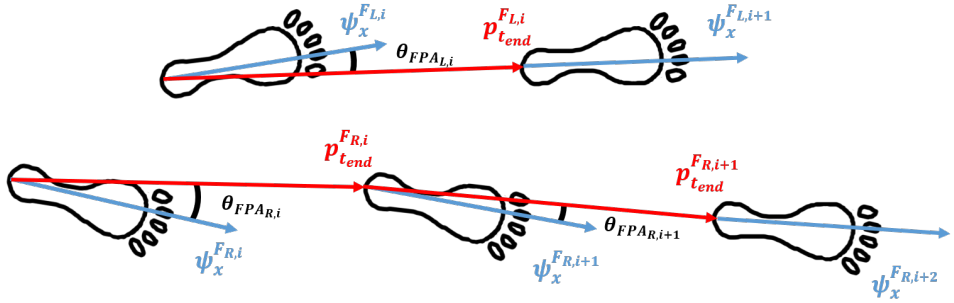


Figure 5.2: A dynamic foot frame, that is initialized in every stance phase i (for left (L) and right (R) separately) and is maintained until the consecutive stance phase of the same foot, is used for calculating the FPA as the angle between the foot direction and the walking direction. The x-direction of this dynamic foot frame ($\psi_x^{F_{L/R,i}}$) aligns with the foot direction. All signals are integrated in this dynamic foot frame to obtain a foot trajectory that ends at the next stance phase. This walking direction is shown in red and defined as the position vector between the calcaneus of two consecutive stance phases (with $p_{t_{end}}^{F_{L/R,i}}$, where t_{end} is the step duration). This allows for direct computation of the FPA.

5.2.1 Stance Phase Detection

During the stance phase there are moments that the foot is approximately still on the ground, hence these moments can be identified using a zero-velocity detection approach [92]. Jimenez et al. developed three conditions for the detection approach, however, this resulted in some cases of short stance phases. Therefore, a fourth condition was included that ensures a minimal length of the stance phase. The following four conditions were used in the current study:

1. The magnitude of acceleration needs to be between 9.0 and 11.0 $\frac{m}{s^2}$ (at time t).

2. The local acceleration variance (σ^2) should be smaller than 0.5 $\frac{m^2}{s^4}$ (in a time period of 0.11 s ($s = 5$ samples), which was experimentally determined) during the stance phase to fulfil this condition, and is defined as:

$$\sigma^2 = \frac{1}{2s+1} \sum_{j=t-s}^{t+s} (a_j - \bar{a}_t)^2 \tag{5.1}$$

   Where the local mean is defined as:

$$\bar{a}_t = \frac{1}{2s+1} \sum_{j=t-s}^{t+s} a_j \tag{5.2}$$

3. The magnitude of the angular velocity should be smaller than 50 $\frac{^\circ}{s}$ (at time t).

4. Stance phase length should be 16 ms at minimum, which ensures that the detection method does not suffer from potential false zero-velocity detections.

5.2.2 Mapping foot frame in sensor frame

A mapping between the sensor frame (ψ^X, red in Figure 5.3) and the foot reference frame (ψ^F, green in Figure 5.3) is required to perform all calculations in ψ^F. This is a fixed rotation, assumed that the sensor does not move relative to the foot. Subjects should perform the following calibration: stand still for 5 seconds, walk 4 steps with a FPA of 0° (i.e., keep foot orientation as straight as possible). The first part of the calibration is used to determine the vertical axis (f_z) of ψ^F using the measured gravitational acceleration. The axis perpendicular to the foot direction (f_y) is determined in the dynamic part of the calibration. Principal Component Analysis (PCA) of the angular velocity is used to determine a common rotation axis (f_y) [93]. The third axis is determined by the cross-product of the other two axes (as it should be perpendicular to both previously defined axes):

$$f_x = f_y \times f_z \tag{5.3}$$

This determines the axis in direction of the foot, i.e., this definition allows for FPA calculation using the angle between heading direction of the step and foot

Figure 5.3: An IMU is secured under the shoelaces and its' coordinate system (ψ^X) is displayed in red. The retroflective markers are places on the second metatarsal and the calcaneus, which is assumed to be a FPA of 0 degrees. The foot reference coordinate system is shown in green (ψ^F).

direction axis. To ensure a proper coordinate system (f_y perpendicular to f_z and thus in the horizontal plane), f_y was subsequently determined by taking the cross-product of f_x and f_z. The mapping of ψ^X to ψ^F can then by performed using the following (constant) rotation matrix (R^{XF}):

$$R^{XF} = \begin{bmatrix} f_x & f_y & f_z \end{bmatrix}^T \tag{5.4}$$

5

5.2.3 Orientation Estimation

Start of a step is defined as the middle of a determined zero-velocity phase (according to the mentioned 4 conditions). The angular velocity is measured in ψ^X, which is rotated to ψ^F by using the determined sensor to foot frame mapping R^{XF} (Figure 5.4A). The dynamic foot reference frame at step i (ψ^{F_i}) is initialized by an identity matrix ($R^{F_i}_{t_0}$), such that the change with respect to this frame can be evaluated using the following differential equation [167]:

$$\dot{R}^F_t = \tilde{\omega}^F R^F_t \tag{5.5}$$

with $\tilde{\omega}^F$ as the skew matrix of the angular velocity, which is defined as:

$$\tilde{\omega}^F = \begin{bmatrix} 0 & -\omega^F_z & \omega^F_y \\ \omega^F_z & 0 & -\omega^F_x \\ -\omega^F_y & \omega^F_x & 0 \end{bmatrix} \tag{5.6}$$

5.2.4 Trajectory Estimation

Figure 5.4B shows the different steps to obtain the foot position ($p^{F_i}_t$). First the measured acceleration (a^X_t) should be transformed to ψ^{F_i} at any time t during step i, such that the gravity component can be subtracted. After this step the acceleration

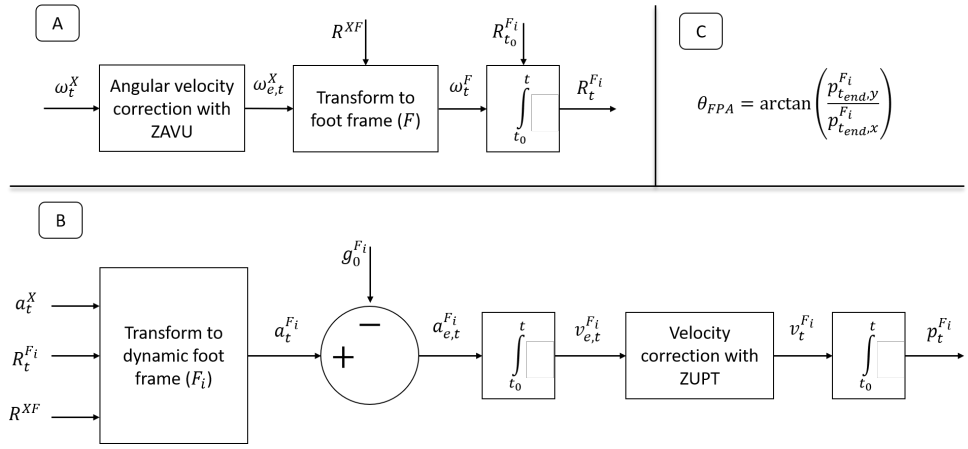


Figure 5.4: FPA estimation approach using gyroscope (ω_t^X) and accelerometer (a_t^X) data: (**A**) sensor angular velocity is corrected using the Zero Angular Velocity Update (ZAVU). Next, it is rotated to the foot frame (ψ^F) by using the mapping between the sensor and foot frames (R^{XF}). The orientation of the dynamic foot frame ($R_t^{F_i}$) is determined by integrating this angular velocity and initializing it with ($R_{t_0}^{F_i} = I$). (**B**) acceleration information is rotated to the dynamic foot frame (ψ^{F_i}), such that the gravitational acceleration can be subtracted to obtain the estimated free acceleration ($a_{e,t}^{F_i}$). This is integrated to velocity ($v_{e,t}^{F_i}$), which in turn is corrected using ZUPT. After another integration step the position of the foot is calculated in the dynamic foot frame. (**C**) Since everything is calculated in ψ^{F_i} the FPA is estimated using a trigonometric relation of the foot position at the end of the step (t_{end}).

is integrated to obtain the velocity (v_t^F) . Since it is known that the velocity should be zero at the next stance phase, we can apply a linear correction (to account for the potential drift) to the velocity vector from start to end of the step. A second integration step is applied to obtain the foot position w.r.t. start of the step (p_t^F).

5.2.5 FPA

The FPA is estimated using the heading vector (endpoint $p_{t_{end}}^{F_i}$ of step i) which is expressed in ψ^{F_i}, therefore the following direct trigonometric relation is applicable here:

$$\theta_{FPA} = \arctan \frac{p_{t_{end},y}^{F_i}}{p_{t_{end},x}^{F_i}} \tag{5.7}$$

5.3 Validation Measurement Protocol

The accuracy of the proposed FPA estimation approach is quantified by comparing results obtained with our approach to those from using an optical motion capture system. Five healthy volunteers (5 males; age: 25.2 ± 4.2 years; height: 1.83 ± 0.09 m; weight: 80.0 ± 9.5 kg; body mass index: 24.1 ± 3.4 kg/m^2) participated in this research in a gait laboratory. All subjects reported no recent injuries that affect

balance or mobility. The ethics committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente approved this protocol and all subjects provided written informed consent prior to the measurements.

Subjects are fitted with one inertial sensor (MTw Awinda, Xsens, Enschede, the Netherlands) on top of the shoe and two retro-reflective markers (placed on the head of the second metatarsal and the calcaneus, as shown in Figure 5.3). The MTw is a wireless inertial sensor that transmits data (at 100 Hz) over Bluetooth, which is recorded using MT Software Suite (Xsens, Enschede, the Netherlands). The position of retro-reflective markers is recorded (at 100 Hz) using eight high-speed infrared cameras (Vicon, Oxford, UK) and processed with Nexus 2.8.2 (Vicon, Oxford, UK). To compare the FPA outcomes of both systems, a synchronization between the inertial and optical systems is required. This is achieved by stamping the right foot at the ground for the start of the measurement. This signal is present in both the optical and inertial measurement data and is used to align both signals. Small misalignments (1-5 ms) are allowable since we are interested in the FPA per different step and not at discrete time indexes.

The reference FPA was determined based on the markers placed on the calcaneus and second metatarsal by calculating the angle between the line connecting both retro-reflective markers and the walking direction vector (defined by a line between calcaneus of the same foot in different stance phases) in the lab reference frame (depends on the camera calibration) [178].

Every measurement started with a 0° FPA calibration (for the inertial approach), which consists of a static part and a dynamic part (as mentioned in Section 5.2.2). Subjects should remain as still as possible with feet pointing forwards (0° FPA) for approximately five seconds, to determine the gravitational axis. Subsequently, subjects were asked to walk with a zero degrees FPA for four steps. A visual reference is provided to subjects by a tape placed on their shoe (shown in Figure 5.3), which shows the foot direction vector. By aligning this with a line on the floor subjects could achieve a FPA close to zero, which was evaluated using the optical reference.

After this calibration trial, subjects were asked to perform 3 sets of 12 trials of walking in a straight line within the measurement volume of the optical motion capture system (10 x 4 m, projected on the floor). Each set of 12 trials consists of walking at their preferred walking speed with either normal, positive (toe-out) or negative (toe-in) FPA. The difference in FPA between each of these 3 walking conditions was self-selected by the subjects.

A difference between FPA estimates for the optical and inertial approach was used for an evaluation of the accuracy of the proposed inertial FPA estimation approach. We decided not to evaluate a root mean squared difference but a mean difference, because the sign of errors is relevant in this situation due to potential spatial misalignment of the zero degrees FPA. After correction for the determined offset, results are presented using a Bland-Altman plot to show the distribution of FPA measured by both the optical and inertial sensing approach [24]. Additionally, a repeated measures one-way ANOVA test [134] is performed to determine if both the inertial and optical approaches can differentiate between the three gait conditions.

5.4 Results

Table 5.1 shows the mean differences (and standard deviation) between the FPA estimated using retro-reflective markers and using the proposed inertial approach. It can be seen that subjects (e.g., S01 and S05) with larger differences (up to 5 degrees) compared to the optical reference also have larger 0° FPA calibration differences.

A more detailed comparison between the FPA (of individual steps) estimated from an optical and inertial approach can be found in Figure 5.5, after correcting for the 0° FPA calibration differences. The correlation between both approaches is shown by the plots on the left (for all three types of gait). Good correlation coefficients ($r^2 > 0.7$) can be observed for all conditions. Furthermore, the mean bias between the inertial approach and the optical reference is small ($< 2.5°$) for all conditions.

Results of a repeated measures one-way ANOVA test between the different gait conditions show that both the inertial approach and the optical reference system can significantly ($p << 0.01$) discriminate between those conditions. These results were obtained for all subjects and both measurement approaches.

Table 5.1: The mean (and standard deviation) FPA differences (in degrees) of the optical approach compared to the inertial approach.

Subjects	Gait			0° FPA
	Normal	Toe-In	Toe-Out	Calibration
S01	-5.22 (±3.35)	-2.39 (±2.61)	-6.58 (±1.72)	-3.99 (±0.37)
S02	0.01 (±2.83)	2.51 (±2.86)	-0.21 (±1.41)	1.38 (±0.75)
S03	1.85 (±4.04)	0.10 (±3.43)	-0.13 (±1.87)	-0.17 (±2.82)
S04	0.20 (±3.43)	0.31 (±3.25)	-1.45 (±2.05)	1.68 (±2.86)
S05	1.69 (±2.67)	5.15 (±1.41)	2.36 (±1.36)	4.71 (±1.14)

5.5 Discussion

The aim of this research was to evaluate an approach for estimating FPA using a single foot-worn inertial (accelerometer and gyroscope) sensor. The proposed approach uses a dynamic step reference frame to calculate the FPA of each step with respect to the foot frame during stance. A comparison with an optical reference shows good correlation and it can effectively differentiate between the different types of gait (normal, toe-in and toe-out).

Table 5.1 shows that an offset between the optical and inertial approach could have impacted the observed differences between both approaches. Such an offset is expected to be the result of a misalignment between the defined 0° FPA for both approaches. To that end, results presented in Figure 5.5 were corrected for the observed differences during the 0° FPA calibration measurement (by adding the mean observed offset in Table 5.1 to the estimated FPA with the inertial approach for each subject individually). This misalignment can occur during the sensor to

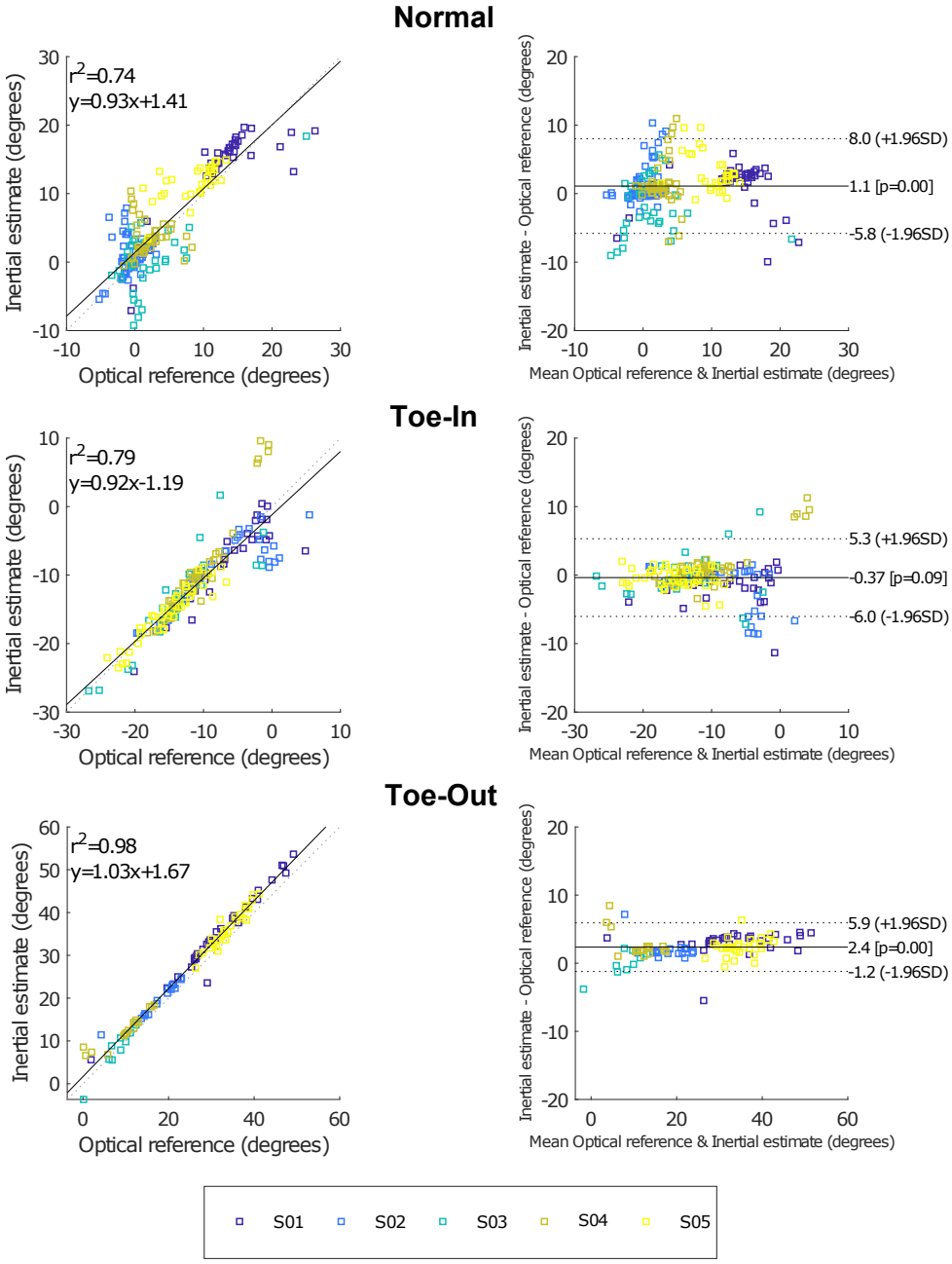


Figure 5.5: Bland-Altman graphs comparing the FPA estimated with an optical and inertial approach for five subjects. The mean observed differences during the 0° FPA calibration trial was added to the inertial outcomes of the individual subjects, such that impact of misalignment of the 0° FPA axes is minimal. Different graphs are presented for the three types of gait (normal, toe-in and toe-out), for each condition approximately 40 steps were analyzed. Please note the differences in angle ranges between the three types of gait.

foot calibration of the inertial approach, since subjects were instructed to walk with an FPA of 0° using optical feedback (tape on the shoe and lines on the floor). Furthermore, misplacement of the retro-reflective markers could also result in an offset between both approaches. The inertial calibration procedure could be improved by using a board with cut-outs for the feet, which forces subjects to walk with 0° FPA. In this manner, a potential misplacement of the retro-reflective markers can also be determined.

Related works of estimating foot angles using inertial sensing reported comparable performance as our proposed approach (maximum mean difference of ∼2.6°). Bidabadi et al. used a single foot-worn IMU to estimate the foot pitch angle (ankle flexion/extension) and reported a mean accuracy of ∼3.8° [175]. Huang et al. presented a single foot-worn IMU (with magnetometer) approach for estimating the FPA with a maximum mean error of ∼2.5° [87]. While a full-body inertial approach for estimating FPA was reported to have an error of ∼2.4° [99]. However, these approaches require more on-body sensors or cannot be used in all (magnetically distorted) measurement environments.

One of the issues with inertial sensing is that directly integrating the accelerometer and gyroscope measurements will result in drift of the sensor position/orientation. However, impact of such drift increases over time, i.e., short-term integration could result in outcomes with acceptable accuracy. To that end, we applied two ways of minimizing such effects, namely ZUPT and integrating over each individual step separately. ZUPT allowed for linear corrections to the obtained velocity/position, due to the known zero-velocity state during stance. And the use a dynamic step frame allows for integration of accelerometer data during each individual step. In this manner, drift only impacts the estimated FPA during a single step, which reduces the negative effect on accuracy substantially.

The proposed approach has potential for real-time feedback applications, such as proposed by Karatsidis et al. [99]. A reduction in the number of sensors is beneficial to the patients, because of the decreased complexity and costs. However, this approach was evaluated with healthy participants with no reported balance or mobility issues. The FPA of people with movement disorders might be estimated with lower accuracy using the proposed approach. Additionally, with different gait dynamics, the zero-velocity detection conditions might change. Furthermore, the calibration procedure (walking with 0° FPA) used in this work might be difficult for people with a movement disorder. An alternative could be to perform repeated dorsal/plantar ankle flexions. However, in initial measurements this resulted in a rotation axis that was not perpendicular to the vertical since the rotation was not consistently in the horizontal plane. If the mapping of sensor to foot frame is known, the calibration procedure might be removed, e.g., in case of a shoe with an embedded IMU [220, 212]. Depending on the application the impact of an incorrect 0° FPA might vary, as long as differences compared to a baseline measurement can be measured with sufficient accuracy [153].

To apply the proposed approach in a (semi-)real-time feedback application a firmware implementation would be required. In the current study, the algorithm was off-line applied in MATLAB, however, minimal calculation time (∼4 ms per step) was observed for this implementation. Furthermore, feedback can only be provided after the step is finished (due to uncertain step direction during swing phase). Therefore, it is expected that this method can provide (semi-)real-time

feedback on the FPA. However, additional research is required to investigate the accuracy of the proposed approach in real-time and with patients.

5.6 Conclusion

This work presented a novel approach to estimate FPA using information from a single foot-worn inertial sensor (accelerometer and gyroscope) that can be used in any (magnetically distorted) environment. Experimental results show that the proposed approach has good correlation with an optical reference system. Furthermore, differences between various types of gait (normal, toe-in and toe-out) can be discriminated with our approach. Therefore, this research could provide a basis for future research into the use of wearable feedback systems for gait training of KOA patients in any environment.

5.7 Acknowledgments

This research (project No. 13917) is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs. The authors would like to thank the Roessingh Research & Development for the availability of the gait laboratory for the measurements.

Part III

Technologies for Motion Capture

Chapter

SIX

On the Validity of Different Motion Capture Technologies For the Analysis of Running

Published as:
F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. J. F. van Beijnum and P. H. Veltink **On the validity of different motion capture technologies for the analysis of running** 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob), Enschede, 2018, pp. 1175-1180 http://doi.org/10.1109/BIOROB.2018.8487210

Abstract

An increasing diversity of available motion capture technologies allows for measurement of human kinematics in various environments. However, little is known about the differences in quality of measured kinematics by such technologies. Therefore, this work presents a comparison between three motion capture approaches, based on inertial-magnetic measurement units (processed with Xsens MVN Analyze) and optical markers (processed using Plug-In Gait and OpenSim Gait2392). It was chosen to evaluate the different motion capture approaches in running, as such kinematics are preferably measured in the natural running environment and involve challenging dynamics. An evaluation was done using data of 8 subjects running on a treadmill at three different speeds, namely 10, 12 and 14 km/h. The sagittal plane results show excellent correlation ($\rho > 0.96$) and RMSDs are smaller than 5 degrees for 6 out of the 8 subjects. However, results in the frontal and transversal planes were less correlated between the different motion capture approaches. This shows that sagittal kinematics can be measured consistently using any of the three analyzed motion capture approaches, but ambiguities exist in the analysis of frontal and transversal planes.

Keywords: Kinematics; Motion Measurement; Optical Sensors; Optical Variables Measurement; Correlation; Protocols; Motion Segmentation.

6.1 Introduction

A wide variety of technologies is currently available for capturing three-dimensional full-body human motion, of which the most established ones are based on video or inertial sensors. Optical systems rely on the idea of tracking retro-reflective markers, placed on pre-defined locations of the different body segments, using infrared cameras [146, 201]. Inertial systems, on the other hand, track the orientation of inertial-magnetic sensors by sensor fusion of the accelerometer, gyroscope and magnetometer output, which can be translated to body segment quantities by using a calibration pose and pre-defined sensor placement on the various body segments [161, 214]. Traditional motion capture was done in a lab via accurate optical systems, while inertial systems were designed to allow for ambulatory measurements, i.e., not restricted to a laboratory setting.

Besides this apparent difference in applicability, an important determinant in the choice for either motion capture technology is the measurement accuracy of parameters of interest. Both inertial-magnetic and optical motion capture accuracy can be influenced by different external factors, e.g., soft-tissue artifacts (STAs) [116], occlusions [41], sensor/marker placement [18, 124], calibration pose accuracy [18, 157] and/or magnetic disturbances [50, 161]. Most of those potential sources of errors can be solved by proper measurement preparation, e.g., camera distribution, marker/sensor placement and clear subject instructions. However, due to differences in sensor/marker placement between protocols and/or technologies, the impact of STAs on the measured kinematics varies [35]. In case of occlusions gap filling can be applied [41], and magnetic disturbances can be mitigated by using advanced sensor fusion techniques [105, 170]. Even though position estimates of retro-reflective markers are highly accurate (due to cameras with high resolution) [126], estimates of relevant quantities at the body level (e.g., body kinematics) might show a degraded accuracy depending on the considered body modeling assumptions [114] (e.g., Plug-In Gait [49, 94], OpenSim [51] or AnyBody [47]).

Performance comparisons, in terms of kinematics, among different motion capture technologies have been reported in several comparison studies. For example, Stief et. al showed that sagittal plane joint angles in gait measured with Plug-In Gait are more accurate than joint angles in other planes compared to their proposed lower-body marker protocol [188]. An inter-laboratory study showed that sagittal kinematics were consistent between labs and protocols [18]. Comparable results were reported in a comparison between different marker protocols, where out-of-sagittal planes joint angles were shown to have lower correlation for the different marker protocols [67]. Ferrari et al. compared two IMMU with optical protocols and concluded that their protocol can potentially be used for clinical gait assessment [68]. Simarly, a comparison between Plug-In Gait and OpenSim showed that an offset was typically observed between both approaches for walking trials [114]. A musculoskeletal model driven by either IMMU or optical motion capture walking data, showed larger correlations for the sagittal kinematics compared to the frontal and transversal kinematics [98]. Dinu et al. compared centre of mass position estimates during stance and jumps based on IMMU and optical motion capture data, and reported that both estimates are similar and could be used interchangeably [57].

However, to the best of our knowledge, no comparison between different motion capture approaches to assess running kinematics has been performed. Therefore,

the aim of this work is to analyze differences in measurement accuracy of different motion capture approaches, namely IMMU, Plug-In Gait and OpenSim. Running on a treadmill was chosen as an evaluation scenario due to constraints that a laboratory setting poses on measuring kinematics of runners in their natural running environment [149] (e.g., on the track, road or in the woods) and the more challenging dynamics involved, compared to gait. Even though IMMUs allow for real world measurements, the analysis was performed on a treadmill in a gait lab to allow for simultaneous measurements with optical and inertial-magnetic motion capture.

6.2 Materials and Methods

In Section 6.2.1 the data collection protocol is introduced. The measurement setup with the different motion capture systems is described in Section 6.2.2. The data processing, including the different motion capture approaches is explained in Section 6.2.3. And the evaluation measures are described in Section 6.2.4.

6.2.1 Data Collection

Experimental data was collected in the gait laboratory of Roessingh Research and Development (Enschede, the Netherlands), with approval from the ethics committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente. Eight healthy subjects volunteered for the study (8 males; age: 25.1 ± 5.2 years; height: 183.7 ± 4.5 cm; weight: 77.7 ± 9.4 kg; body mass index: 23.0 ± 2.5 kg/m^2). The runners were recruited from a local track & field club and had no recent (self-reported) history of injuries. After a warm-up session of approximately three minutes at self-selected running speeds, subjects were asked to run for three minutes at three different speeds (namely 10, 12 and 14 km/h, in this order) on an instrumented treadmill.

6.2.2 Measurement Setup

Subjects were equipped with both inertial-magnetic and optical motion capture technologies and they were running on an instrumented treadmill (S-Mill, ForceLink, Culemborg, the Netherlands), with a running area of 250 x 100 cm. Xsens MVN Link (Xsens, Enschede, the Netherlands) was used to capture full-body movements at 240 Hz using 17 inertial-magnetic sensors, which were placed in a full-body Lycra suit (as shown in Fig. 6.1). The suit was adapted in such a way to accommodate placement of 41 retro-reflective markers on the subject's skin. The position of these markers, placed according to the Plug-in Gait protocol[1] (Nexus 1.8.5, Vicon, Oxford, UK), was recorded at 100 Hz by six high-speed infrared cameras (MX-13, Vicon, Oxford, UK).

6.2.3 Data Processing

Only periods of steady state running at a constant speed were included in the analysis. Optical motion capture data (recorded with Vicon) was processed using

[1] To allow for OpenSim processing, one additional marker was placed on all lower-body segments.

Figure 6.1: Placement of the 41 retroflective markers (according to the Plug-In Gait protocol)1 can be seen for a representative subject. The markers were placed directly on the subject's skin. IMMU sensors were placed inside a Lycra suit, such that these remain in a fixed position with respect to the body. To accommodate the retroflective markers, holes were cut in this suit. This allowed for simultaneous measurement of kinematics using optical and IMMU motion capture technologies.

both Plug-In Gait (a direct kinematics method) [49, 94] and OpenSim (Gait2392, an inverse kinematics method) [51], which for simplicity will be referred to as PiG and OS, respectively. IMMU sensor data was processed using MVN Analyze (Xsens, Enschede, the Netherlands), referred to as MVN. These motion capture approaches have been used for analysis of running kinematics in other studies [60, 79, 109, 149, 155, 189]. Kinematic data was assessed in the joint frame, and all approaches use Euler angles in the order of flexion/extension (F/E), ab-/adduction (A/A) and internal/external rotation (I/E). For both OS and PiG a static trial was recorded, where the subject performed a neutral pose (standing upright with arms straight next to his/her side) for approximately 5 seconds. The MVN calibration required the same neutral pose, followed by a short walk [170].

Even though full-body kinematics were measured, the analysis was limited to the lower-body outcomes since these are most important in evaluating running kinematics. OS assumes a limited number of degrees of freedom in the knee and ankle joints (modeled as hinges, i.e., only F/E movement is allowed), the knee model of OS was adapted to also allow for A/A and I/E, similar to [142, 170]. In this manner the rotational axes of the adapted OS knee joints aligned with both the PiG and MVN knee joint axes. The transversal and frontal ankle degrees of freedom are modeled in the subtalar joint for OS, and are therefore not modified as these do not align with the ankle joint axes for the PiG and MVN outcomes. All three rotational degrees of freedom are allowed for the lower-body joints for the PiG and MVN models.

For a fair comparison, all data was resampled at 120 Hz using linear interpolation (upsampling) for the optical data and decimation (downsampling) for the IMMU data. MATLAB R2017a (Mathworks, Inc., Natick, MA, USA) was used for processing all data.

6.2.4 Evaluation Measures

Pearson's correlation coefficients (ρ) and Root Mean Squared Differences (RMSD) were used to characterize differences in the measured lower-body kinematics of PiG, OS and MVN. This analysis was performed for joint angles in all three planes: sagittal (F/E), frontal (A/A) and transversal (I/E). Segmentation of the kinematic data was required to allow for analysis of individual strides. This segmentation was done by identifying heel strike events using force output of the instrumented treadmill. Mean correlation over all strides was calculated using a Fisher transformation [45]. Possible offsets between the motion capture approaches were evaluated, in which the offset was defined as the mean difference between both outcomes. Furthermore, two relevant discrete outcome measures for running analysis were compared for the different approaches, namely maximum knee F/E angle during stance and the ankle F/E angle at heel strike [30, 149].

6.3 Results

Results of the comparison between the three different motion capture approaches are shown in Table 6.1, which is divided according to the planes of movement, namely F/E (top), A/A (middle) and I/E (bottom). This table displays the mean correlation coefficient, RMSD (mean and standard deviation over the different strides) and (possible) offset of the measured running kinematics at 12 km/h. Similar results were obtained for other speeds, but are not shown here due to space limitations. Fig. 6.2 shows the mean (and standard deviation) joint angle profiles of two representative subjects when running at 12 and 14 km/h (shown in Fig. 6.2A and 6.2B, respectively).

An excellent correlation ($\rho > 0.96$) between the different approaches was found for all the measured sagittal plane lower-body joint angles (Table 6.1, F/E of the hip, knee and ankle). This can also be seen in Fig. 6.2, where the largest difference is found in the hip F/E angles.

Fig. 6.2 shows MVN, PiG and OS give inconsistent joint angle profiles in the frontal (Table 6.1, A/A of the hip, knee and ankle) and the transversal (Table 6.1, I/E of the hip, knee and ankle) planes. The hip A/A angles show largest correlations compared to other joints/planes, where the RMSDs were not larger than 8 degrees between the different approaches. For other joints RMSDs up to 30 degrees were observed, which in some cases (e.g., ankle A/A and I/E) can be largely contributed to an offset between the different approaches.

Table 6.1: Individual results (evaluated at 12 km/h, for all three rotational axes: F/E, A/A and I/E) are compared between the different motion capture approaches: MVN, PiG and OS using ρ, RMSD (σ) and offset. Note that ankle A/A and I/E are missing for OS since these are not assessed in the model.

Kinematics 1 Kinematics 2		MVN PiG			PiG OS			OS MVN		
Flexion/extension angles										
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Hip	S01	0.994	8.86 (0.76)	-8.44	0.991	2.64 (0.39)	-0.82	0.986	9.84 (0.80)	9.26
	S02	0.995	11.97 (0.78)	-11.47	0.989	4.13 (0.35)	-2.64	0.994	14.52 (0.45)	14.11
	S03	0.993	4.02 (0.80)	-1.02	0.989	4.83 (1.06)	-3.91	0.992	5.60 (1.11)	4.93
	S04	0.990	8.81 (0.55)	-8.41	0.990	7.24 (1.04)	-6.76	0.992	15.38 (1.02)	15.16
	S05	0.986	8.91 (1.67)	-8.24	0.992	5.87 (0.61)	-5.31	0.985	13.99 (0.96)	13.55
	S06	0.992	7.11 (1.02)	-6.75	0.991	6.66 (0.63)	-6.20	0.995	13.16 (0.66)	12.95
	S07	0.996	12.05 (0.93)	-11.82	0.990	6.91 (0.69)	-6.26	0.988	18.35 (0.37)	18.08
	S08	0.990	15.30 (1.06)	-15.06	0.990	2.97 (0.68)	-1.70	0.984	17.03 (0.49)	16.76
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Knee	S01	0.992	7.85 (1.31)	6.83	0.988	7.21 (3.93)	-5.61	0.985	6.29 (0.63)	-1.22
	S02	0.992	7.57 (2.82)	5.56	0.986	7.41 (3.98)	-5.49	0.988	6.86 (0.79)	-0.07
	S03	0.986	9.93 (3.76)	7.65	0.985	7.09 (4.27)	-5.48	0.984	7.15 (1.07)	-2.18
	S04	0.983	9.59 (1.68)	6.84	0.988	10.78 (5.10)	-10.18	0.985	6.81 (1.58)	3.34
	S05	0.981	10.18 (3.04)	7.31	0.982	8.70 (5.50)	-7.90	0.979	7.99 (1.33)	0.58
	S06	0.990	8.71 (1.71)	7.00	0.976	9.65 (9.51)	-7.33	0.964	8.56 (8.74)	0.33
	S07	0.992	6.65 (1.77)	4.20	0.985	7.48 (5.47)	-4.31	0.987	6.39 (4.30)	0.12
	S08	0.982	8.84 (2.50)	5.87	0.988	8.26 (4.34)	-7.42	0.980	7.28 (1.85)	1.55
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Ankle	S01	0.984	3.85 (1.10)	-2.49	0.961	3.99 (0.92)	1.15	0.971	4.50 (1.32)	1.35
	S02	0.975	4.00 (0.73)	-0.85	0.967	6.01 (0.25)	4.62	0.978	5.28 (0.64)	-3.77
	S03	0.966	4.54 (1.55)	-2.96	0.974	6.31 (2.61)	5.24	0.986	3.41 (1.19)	-2.28
	S04	0.981	6.87 (1.32)	-5.88	0.968	4.75 (1.04)	2.38	0.986	4.39 (0.58)	3.50
	S05	0.984	4.68 (1.22)	-3.74	0.960	4.98 (0.16)	3.13	0.983	3.21 (0.83)	0.61
	S06	0.975	6.24 (1.45)	-5.42	0.968	6.49 (0.43)	5.57	0.976	3.03 (0.50)	-0.16
	S07	0.988	8.06 (1.52)	-7.58	0.969	6.79 (0.62)	5.65	0.983	3.65 (1.20)	1.93
	S08	0.984	3.86 (1.06)	-2.62	0.970	4.55 (0.18)	3.14	0.989	2.80 (0.33)	-0.52
Ab-/adduction angles										
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Hip	S01	0.823	4.25 (0.96)	0.10	0.761	8.38 (4.32)	-3.50	0.904	2.98 (0.29)	1.92
	S02	0.757	5.44 (3.35)	-2.04	0.483	5.70 (1.23)	-1.04	0.633	6.74 (0.43)	-0.05
	S03	0.683	2.76 (0.89)	-0.63	0.587	3.98 (1.05)	-0.63	0.467	3.36 (0.31)	0.92
	S04	0.672	4.72 (0.44)	0.60	0.689	7.32 (1.06)	4.13	0.747	3.79 (0.65)	2.58
	S05	0.391	6.06 (0.87)	0.52	0.868	5.24 (2.84)	0.13	0.568	3.84 (0.28)	1.52
	S06	0.772	4.04 (1.15)	-2.95	0.814	5.98 (2.05)	-2.05	0.935	3.79 (2.04)	3.22
	S07	0.666	5.10 (0.28)	0.28	0.743	5.80 (2.45)	-2.45	0.837	4.32 (0.87)	3.66
	S08	0.299	4.84 (1.13)	1.13	0.640	5.18 (1.80)	-1.80	0.765	3.27 (0.29)	1.78
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Knee	S01	0.637	4.77 (1.31)	2.76	0.341	10.77 (3.26)	-6.51	0.072	10.33 (3.92)	4.15
	S02	0.876	4.11 (2.82)	-2.32	0.511	12.93 (5.52)	-2.84	0.392	10.40 (4.37)	0.08
	S03	0.708	16.44 (3.76)	-8.83	0.484	17.91 (0.94)	0.03	0.351	13.01 (1.61)	0.79
	S04	0.372	8.89 (1.68)	-6.66	0.542	12.81 (6.20)	-8.70	0.312	12.90 (0.63)	-10.09
	S05	0.695	6.56 (3.04)	2.90	0.546	10.84 (4.72)	-5.78	0.398	12.46 (1.80)	8.69
	S06	0.244	8.29 (1.75)	-3.14	0.503	16.90 (5.57)	-13.55	0.228	11.43 (2.44)	0.16
	S07	0.532	11.67 (1.77)	-7.85	0.637	12.16 (6.64)	0.31	0.500	11.18 (3.64)	4.89
	S08	0.347	8.03 (2.50)	-3.66	0.150	10.65 (1.10)	-1.00	0.340	11.32 (0.78)	8.06
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Ankle	S01	0.716	9.75 (1.28)	6.50						
	S02	0.518	29.21 (6.65)	28.04						
	S03	0.591	30.47 (11.04)	29.81						
	S04	0.585	12.54 (4.97)	5.13						
	S05	0.649	11.39 (2.30)	1.71						
	S06	0.467	11.98 (0.64)	-2.38						
	S07	0.690	16.56 (4.50)	12.27						
	S08	0.595	14.55 (6.20)	12.82						
Internal/external rotation angles										
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Hip	S01	0.287	8.25 (0.67)	-3.13	0.248	13.87 (2.70)	-5.96	0.213	12.83 (3.72)	0.46
	S02	0.119	12.39 (6.54)	10.05	0.349	14.80 (2.49)	3.99	0.166	16.78 (2.09)	-6.07
	S03	0.134	15.16 (8.21)	-9.46	0.483	20.77 (6.31)	12.59	0.102	13.28 (3.73)	-3.13
	S04	0.236	19.24 (9.73)	13.31	0.722	19.61 (7.08)	8.56	0.139	13.26 (3.34)	-1.23
	S05	0.442	11.47 (1.30)	-0.90	0.566	14.70 (3.69)	-2.48	0.535	14.19 (2.41)	3.37
	S06	0.555	23.76 (4.84)	-22.96	0.197	29.89 (5.42)	-26.31	0.427	14.45 (4.74)	3.37
	S07	0.042	33.36 (3.67)	3.66	0.366	39.71 (0.66)	0.43	0.177	16.98 (0.91)	8.03
	S08	0.369	12.59 (6.90)	-9.91	0.177	14.78 (1.54)	7.36	0.456	12.88 (1.97)	6.86
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Knee	S01	0.142	4.77 (1.18)	-3.49	0.522	13.38 (3.26)	9.37	0.167	14.17 (3.92)	3.09
	S02	0.392	4.11 (0.95)	-2.65	0.572	22.22 (5.52)	18.51	0.091	17.41 (4.37)	1.78
	S03	0.237	16.44 (9.48)	-15.58	0.327	22.31 (0.94)	17.05	0.372	13.51 (1.61)	-1.48
	S04	0.312	8.89 (1.01)	-10.40	0.514	19.37 (6.20)	7.78	0.460	16.93 (0.63)	2.62
	S05	0.412	6.56 (3.96)	0.40	0.546	23.07 (4.72)	1.89	0.550	16.99 (1.80)	-2.29
	S06	0.676	8.29 (3.26)	-16.43	0.630	19.73 (5.57)	15.87	0.352	16.62 (2.44)	0.56
	S07	0.220	11.68 (5.85)	7.29	0.697	29.30 (6.64)	26.60	0.365	17.37 (3.64)	-4.38
	S08	0.446	8.03 (4.52)	2.15	0.490	17.29 (1.10)	13.40	0.441	12.27 (0.78)	-1.77
	Subjects	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset	ρ	RMSD (σ)	Offset
Ankle	S01	0.085	10.94 (0.50)	-9.59						
	S02	0.419	9.17 (1.89)	-8.44						
	S03	0.143	13.45 (8.85)	-9.11						
	S04	0.396	9.07 (2.30)	7.73						
	S05	0.154	9.56 (5.26)	-6.79						
	S06	0.555	4.11 (0.62)	-0.75						
	S07	0.310	7.24 (0.70)	0.17						
	S08	0.426	5.63 (0.96)	-4.63						

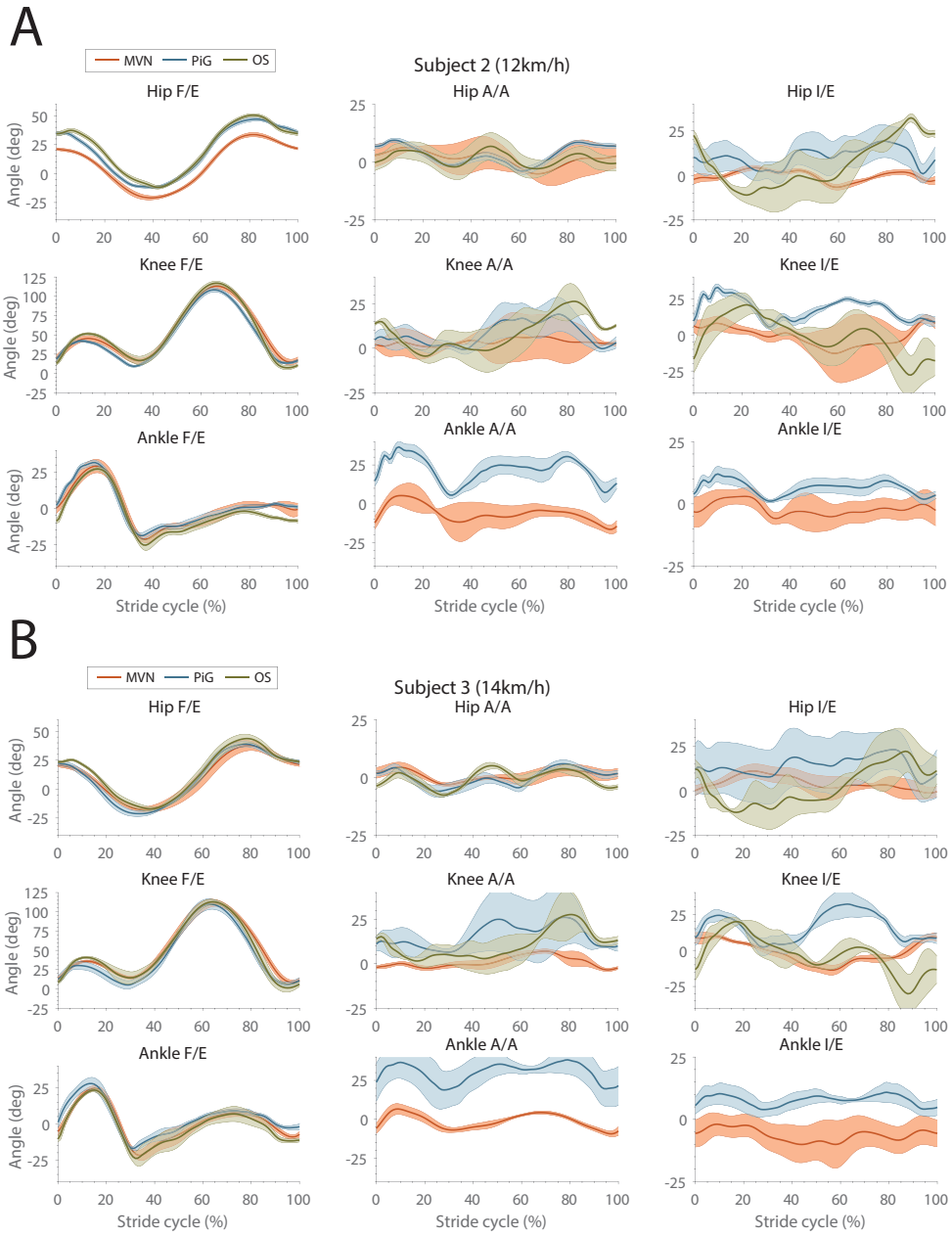


Figure 6.2: Comparison between running kinematics of two representative subjects (**A** shows subject 2 at 12 km/h and **B** shows subject 3 at 14 km/h) obtained from three motion capture approaches: MVN (orange), PiG (blue) and OS (green). Angles are displayed as a mean (and standard deviation band) normalized to the stride cycle. The left column shows flexion/extension (F/E), in the middle ab-/adduction (A/A) and the right column internal/external (I/E) rotation for the hip, knee and ankle joints. These results are obtained from approximately 400 strides. Corresponding correlations, RMSD (standard deviation) and offset can be found in Table 6.1.

Boxplots of the measured maximum knee F/E angle during stance and the ankle F/E angle at heel strike (Fig. 6.3) show differences of up to 10 degrees between the mean outcomes of the different approaches. Fig. 6.3A shows that relative differences in the maximum knee F/E angle between different approaches are consistent over the various subjects, as in all cases OS shows the largest and MVN the smallest outcomes. Similarly, Fig. 6.3B shows largest mean ankle F/E angles at heel strike for PiG and smallest for OS in the majority of the subjects.

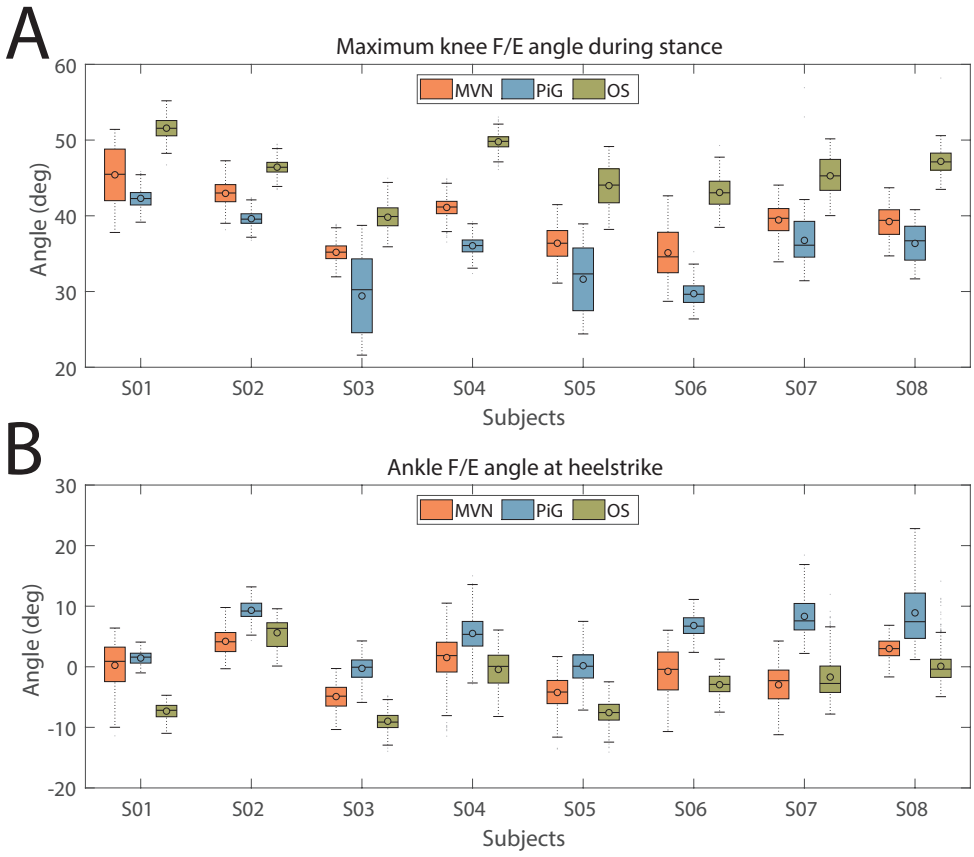


Figure 6.3: Comparison between discrete running outcomes (at 12 km/h) of all measured subjects from three motion capture approaches: MVN (orange), PiG (blue) and OS (green). **A** shows the maximum knee F/E angle during stance, and **B** shows the ankle F/E angle at heel strike. These box plots were obtained from approximately 400 strides.

6.4 Discussion

The results show excellent correlations ($\rho > 0.96$) for the sagittal plane joint angles (Table 6.1, top part), while the frontal and transversal planes show larger differences between the different motion capture approaches, also when based on the same technology (optical motion capture, processed with PiG and OS). This is a relevant outcome as for both the knee and ankle joints the main axes of movement is F/E

during running; however, other axes are also considered important for assessment of running kinematics [136]. Soft-tissue artifacts (STAs) can be a source of observed differences between the different technologies (optical and inertial-magnetic motion capture), but not for the differences in kinematics obtained by processing with PiG or OS [35]. It has been shown that STA have the smallest impact in the sagittal plane [186]; however, such STA effects can be further reduced by combining video with IMMU information [23]. Another factor mostly impacting non-sagittal planes is the potential misalignment of actual joint axes with the ones in the biomechanical model [20]. This misalignment then results in interference between the different axes of rotation. A dynamic calibration can minimize such misalignment effect, as was performed for the IMMU processing [170].

Table 6.1 reports offsets for most of the joints; however, it should be noted that when correlations are weak, RMSD between the approaches cannot be solely attributed to the calculated offsets. On the other hand, the hip F/E offset can be mostly contributed to an offset between different approaches (MVN, PiG and OS) as it is of similar size as the RMSD and correlations are excellent ($\rho > 0.98$). MVN shows the largest offset compared to both PiG and OS, since this offset is observed consistently for all measured subjects, it is likely caused by the calibration procedure [18, 157]. Pelvic tilt can differ between subjects in the calibration pose while identical assumptions about that pose are still applied to each subject. A knee F/E angle offset of 4.8 degrees was reported by Lathrop et al. between OS and PiG, and similar results (6.7 degrees on average) were found in this work as shown in Table 6.1 [114].

Observed differences in ankle F/E angles at heel strike (Fig. 6.3B) could result in inconsistent categorization of runner's phenotypes based on their landing pattern [30]. More generally, this may translate to interpretation of general gait classifications, such as disease severity [69]. Due to differences of up to 10 degrees for both discrete outcomes it is difficult to compare such outcomes fairly between the different approaches. However, the observed offsets (Table 6.1) in sagittal knee and ankle joints could have contributed to these differences in maximum knee F/E during stance and ankle F/E angle at heel strike. Furthermore, due to the consistency of relative differences over the various measured subjects, such differences between approaches can potentially be compensated for.

Even though differences in the frontal and transversal planes are larger than in the sagittal plane, observed (in Fig. 6.2) joint angle profiles in those planes were similar over the various subjects and speeds within each motion capture approach. To assess consistency of the frontal and transversal planes more work should be done to also take into account marker placement variability. Therefore, comparing motion capture data from various approaches in non-sagittal planes can result in inconsistent interpretations and conclusions about the measured kinematics.

6.5 Conclusions

In this work a comparison between three different motion capture protocols was presented for the analysis of running kinematics, namely based on IMMUs (MVN) and optical markers (PiG and OS). Joint angles in the sagittal plane were measured with the largest similarity in terms of high correlation coefficients between all three

approaches. However, differences between the approaches existed and were larger in the frontal and transversal planes than for the sagittal plane, which could lead to inconsistent interpretations. Unlike general opinion which labels optical systems as the golden standard for motion capture, this work shows that clear disagreement between optical kinematic data processed with PiG and OS is present in the frontal and transversal planes, while consistent temporal behavior patterns were found in the sagittal plane. Therefore, the choice of one motion capture approach over the other ones will impact results and may impact interpretation of the measured phenomena. And hence given the importance of measuring running in a natural environment, using an IMMU system can be an equally reasonable alternative to optical motion capture systems for sagittal analyses.

6.6 Acknowledgements

The authors would like to thank Roessingh Research & Development (Leendert Schaake in particular) for providing access to the gait laboratory and help with processing the optical motion capture data.

Chapter

SEVEN

General Discussion and Conclusions

7.1 Overall Insights

In this thesis a number of approaches for minimizing the sensor set for full-body motion capture have been developed and evaluated. The first two studies focus on the development of machine learning methods to achieve full-body motion capture with a minimal inertial sensor set. The third study applies this knowledge to an application for estimation of running kinematics and kinetics. The fourth study applies mechanical modeling to estimate foot progression angles using a single foot-worn inertial sensor by using a dynamic reference frame. The fifth study compares the kinematic outcomes obtained from inertial and optical motion capture technologies.

The first objective of this thesis is the development, optimization and evaluation of methods for estimating full-body poses using a minimal sensor set. This objective is divided into two sub-questions. Firstly, what is the difference in performance for using either lazy or eager learning for estimating full-body poses using a minimal inertial sensor set? This is investigated in Chapter 2. Results in this study show that both learning paradigms can be applied effectively to reduce the number of sensors in full-body motion capture. However, jitter was present in the output poses, because of the lack of time coherence between different poses. This led to a second sub-question, namely: how can knowledge of time coherence be applied to estimation of body poses over time? To that end, in Chapter 3 a stacked input neural network was proposed, that uses information from adjacent (in time) poses as input to the network, which was compared to the current state-of-the-art deep learning approach for estimating full-body poses using a minimal sensor set. We found that comparable accuracy can be achieved when shallow learning is combined with stacked inputs instead of using a deep learning approach. Therefore, this shallow learning approach provides an interesting alternative to deep learning ones, because it requires less training data and computational resources. It should be noted that higher jerk was observed, which was likely the consequence of not explicitly enforcing time coherence (requiring relations between poses in time) as is done in recurrent neural networks.

The second objective of this thesis is to develop applications that use minimal sensing for pose estimation, which led to the development of three applications. Firstly, methods evaluated in the first objective have been applied to estimate sagittal knee kinematics and vertical ground reaction forces during running with three inertial sensors in Chapter 4. In this study, it was shown how fine-tuning of the developed methods to a specific application can increase the estimation accuracy significantly. Additionally, these methods have been extended to also estimate kinetics based on a small inertial sensor set. This is an important step to move running analysis out of the lab, since this is typically done using optical motion capture in combination with force plates. Chapter 5 describes an application for estimation of the foot progression angle using a single foot-worn inertial sensor without magnetometer. We chose to use an approach driven by mechanical principles to estimate these angles, since proper assumptions were sufficient for reducing the number of sensors. Hence no training of data-driven models was required, but a calibration procedure is sufficient. The use of a dynamical foot reference frame was shown to provide a drift-free estimate of the foot progression angle, because all signals are integrated during a single step only. The results show comparable performance to a method that uses the magnetometer, while our proposed method can be applied in any (magnetically distorted) environment. Preliminary work showed that using a

magnetic sensing system for estimation of end-effector position has potential, which is described in Appendix A. This is a different method of applying minimal sensing compared to measuring body segment orientations using inertial sensors (combined with body segment length information) on each intermediate body segment to calculate the end-effector position. However, more research is required to evaluate the applicability of this technology for use in a VR training application for stroke survivors.

All previous studies involved the use of inertial motion capture, however, the current golden standard is optical motion capture. To that end, the measured kinematics from both inertial and optical motion capture systems are compared in Chapter 6. The results show that sagittal joint kinematics are comparable between both systems, while differences in the frontal and transverse planes are larger, even for different accepted processing methods of the same optical data. These differences are likely the result of different soft-tissue artifacts or joint axes definitions between the used technologies. However, the results were consistent within each method. Therefore, comparing kinematics using the same methods can be done fairly, while comparing over different methods might lead to inconsistent conclusions.

The studies described in this thesis show the substantial potential for reducing the number of sensors in full-body motion capture, as illustrated by the different application studies. This may enable applications during daily-life, in which unobtrusiveness of the applied sensing modalities is essential. In the next sections, we discuss directions for future research, such as the use of machine learning or mechanical modeling, deep or shallow learning, the opportunities and limitations of the increasing amount of data, application possibilities and differences between optical and inertial motion capture. These topics are important to further drive development of approaches for reducing the number of sensors for full-body motion capture.

7.2 Machine Learning versus Mechanical Modeling

Methods presented in Chapters 2, 3 and 4 are all based on machine learning approaches. The learned models from such data-driven methods can be rather difficult to comprehend, as these are regarded as black-box methods. For neural networks such insight could be obtained when handling images, as a visual representation handled by the network can be shown [216]. This is difficult in the case of orientational input data, which might not have a physical interpretation in the different hidden layers of a neural network. The field of explainable artificial intelligence could provide such insights regardless of the input features [1], therefore, this direction is an interesting step towards more understandable machine learning methods. Mechanical modeling, using mechanics of a particular model with appropriate assumptions as was applied in Chapter 5, on the other hand does not lack such insights and does not require training based on a dataset. This approach used mechanical principles to estimate foot progression angles with a single foot-worn sensor, because complexity of this problem could be sufficiently contained in a simple mechanical model of foot movement and inertial sensing during these movements. Such an approach, based on physical understanding may, however, not always be feasible. Both approaches have their (dis)advantages, mechanical approaches require highly specific

models to reduce the number of sensors, e.g., [87, 149], while machine learning tools are capable of modeling such complex relations based on datasets in a more general sense [4]. Choosing either approach depends on a good analysis of the problem at hand. Mechanical modeling may be the choice if the specific problem can be solved adequately based on mechanical principles. If a more general approach is needed, machine learning may be the best choice, requiring training and an appropriate dataset, but this should preferably be applied with good understanding of the specific problem. Therefore, future research in reducing the number of sensors for full-body motion capture should focus on combining the best of both approaches, as demonstrated by the promising results when physical body constraints are applied to aid the training of human body poses [48, 88, 219].

7.3 Deep versus Shallow Learning

Deep learning is gaining more and more attention in the field of machine learning [117], since the vanishing gradient problem[1] has been resolved by improved network architectures [83]. In deeper neural networks feature engineering is less important, as higher level features are constructed by the first few hidden layers [117]. This has driven most recent developments in speech [77], image [181] and video recognition [215]. However, training such deep neural networks requires significant computing power and large datasets. As an alternative a shallow network for pose estimation using stacked inputs (implicitly modeling recurrency of the inputs) was evaluated in Chapter 3. In this manner, the benefits from a recurrent deep learning approach can be applied to a shallow approach. This resulted in comparable joint positions errors for estimating full-body poses using either approach. However, increased jerk errors for the shallow approach compared to the deep learning one were observed, which indicates that explicit recurrency modeling can provide more smooth outcomes. An advantage of using a shallow learning approach is that it can be adopted to a firmware implementation, which allows for applications of minimal sensing with only on-body sensors and processing.

7.3.1 Dataset Possibilities and Limitations

The increasing availability of public human motion capture datasets (e.g., CMU [33], AMASS [122] and HDM [129]) provides great opportunities for training deep learning approaches. However, composition of these datasets is rather limited, because most data is recorded using only optical motion capture and movements typically include gait, running, specific sports and ADL. Examples of data that could extend such datasets are: movements of people with gait disabilities, a larger variety of sports or movements in working environments, etc. For applications that require different data than can be obtained directly from optical systems, e.g., accelerations, only limited amount of data is available. Such quantities can be simulated using the optical data, but this will require significant effort in cleaning up such derived data [88, 202]. However, differences in simulated and measured inertial data could be large,

[1] During training of a neural network weights of the individual neurons are adapted proportional to the partial derivative of the error function. In case of a deep neural network this derivative might approach zero preventing the weights to be adapted and might result in a halt of training.

for example due to soft-tissue artifacts which impacts both technologies differently because of variations in sensor and marker placement. This can result in ineffective training of a neural network based on simulated inertial data.

When the application involves movements not included in such datasets, e.g., golfing, additional motion capture data is likely required to reach the desired accuracy. For this reason, an inertial motion capture dataset that contains a wide variety of movements was collected for evaluating the proposed methods in Chapters 2, 3 and 4. However, this dataset is still limited in variation of movements, age and health of subjects. Therefore, it is recommended to extend public databases with more variety in the used technologies and recorded movements to drive innovation in reducing the number of sensors for full-body motion capture.

Another approach would be to leverage knowledge obtained in a similar field, namely pose estimation from images, as large datasets containing a wide variety of poses are available, e.g., MPII [131], SURREAL [200], UP-3D [113] and DensePose [78]. These large datasets allow for adequate training of deep neural networks for estimation of poses from two-dimensional images [10, 193]. Which provides opportunities to apply transfer learning[2] to trained networks for image pose estimation and transfer that knowledge to pose estimation using a minimal body-worn sensor set [139]. This would allow to benefit from these large publicly available datasets of images with human poses and be less limited by the smaller datasets of three-dimensional motion capture data. However, this would not be straightforward, since the features of both inputs are in different domains, i.e., image and sensor data. Furthermore, this technique could likely be applied to use the optical motion capture data to train neural networks and transfer that to inertial motion capture data inputs, because both data types are closely related.

Besides transferring knowledge from one data domain to another (e.g., 2D image to 3D motion capture), it would be beneficial to extend the currently available motion capture datasets with wider ranges of motions and used sensor technologies. When that happens it could result in an increase of the development of methods for reducing the number of sensors, similar to what has happened to image classification with the release of ImageNet [52]. Such developments could really take reduction of the number of sensors for full-body motion capture to the next level.

7.4 Motion capture technology

Various examples of different technologies for full-body motion capture using minimal sensing can be found, namely magnetic sensing [14], ultra-sound [101], pressure plates [96], optical trackers [36] and IMUs [88, 120, 152, 202]. This shows that similar techniques can generalize over sensor technologies. Throughout this thesis we have mainly focused on the use of inertial sensors (Chapters 2, 3, 4 and 5), because no external infrastructure is required and can therefore be used in various environments. However, the choice for such sensor technology should largely be driven by a specific application and its outcome measures of interest (e.g., positions, accelerations or orientations), since differences in measurement accuracy of such quantities exist between the distinct technologies. For example, optical technology measures

[2]Using knowledge obtained from training on one problem and applying it to a different (but related) problem.

marker position with millimeter accuracy [64]. While an IMU based system can achieve an orientation accuracy of approximately 0.5 degrees [119]. Additionally, accelerometers essentially measure the force acting on a mass, of which the acceleration is calculated using Newton's law. Therefore, inertial sensors are better suited for kinetic analyses [39, 61, 133]. Furthermore, a magnetic measurement system can achieve similar position accuracy as optical systems [106, 159, 160, 169].

Inertial motion capture is used in various studies of reducing the number of sensors for full-body motion capture [9, 88, 185, 192, 202], which is likely because of the ease of setting up measurements and can be used in any environment. As this was the reason for using mainly inertial motion capture data in this work. However, it was also shown that our developed methods are not only limited to this type of technology. As was shown in Chapter 4 where our proposed methods were also trained using optical motion capture data. In general, this optical technology is applied to measure full-body kinematics in a laboratory at the level of positions/orientations [206]. However, in Chapter 6 it was shown that differences in the transversal and frontal planes among processing techniques of optical data exists, this indicates that kinematics measured with different motion capture technologies or processing techniques cannot be compared directly. It should be noted that the sagittal kinematics are comparable for all different evaluated technologies and processing techniques, which is likely because this is the principal movement direction of many joints in the human body. Furthermore, accuracy of inertial technology is constantly improving [170]. Therefore, inertial motion capture is a good alternative to optical technologies, that allows for a wide range of applications. However, this would require additional comparison research for various measurement scenarios, e.g., stroke patients or virtual reality.

7.4.1 Sensor Locations

Regardless of the choice for a type of motion capture, the required on-body sensor placement is important for the resulting accuracy of estimating full-body poses using a minimal sensor set. Sensor locations have been investigated in Chapter 2, where the optimal sensor locations were determined by training different neural networks with a small set of possible sensor locations and select a configuration with the smallest error. This configuration might not be the best choice for all applications due to application constraints or different motion scenarios. Therefore, a more generic approach to determine this optimum could provide a significant reduction of effort for the development of specific applications with minimal sensing. Information theory has the potential to provide such generic tools, which has been applied to feature engineering for activity recognition [110]. Within the field of information theory, mutual information has the largest potential for indicating the most informative segments, because it provides insight in segments that have overlapping information [138, 137]. Alternatively, principal component analysis could provide similar insights in the segments that describe a pose sufficiently [21].

7.5 Application Possibilities

In Chapters 4 and 5 and Appendix A, minimal sensing was applied to specific applications. However, there are many more application fields that could benefit from reducing the number of sensors for full-body motion capture. The largest application possibilities are in cases that require truly ambulatory measurements and/or involve users that cannot easily apply a large number of sensors or markers, e.g., patients and athletes. Furthermore, reducing the number of sensors/markers results in cheaper motion capture solutions, due to the decreased amount of required hardware. It should be noted that increased computational complexity might impact this economical advantage, which is minimal with the proposed stacked input neural network. Examples of applications that benefit from proposed methods in this thesis are (but are not limited to): real-time feedback on the foot progression angle [99], virtual reality training of stroke survivors [115] and assessing daily-life reaching performance of stroke survivors [199].

7.6 Concluding remarks

In this thesis it was shown that data-driven methods have potential to reduce the number of sensors in full-body motion capture (Chapters 2 and 3). Additionally, this methodology has been applied to estimate running kinematics and kinetics (Chapter 4). With a mechanical-based model it was shown that the foot progression angle can be accurately estimated with a single inertial sensor (Chapter 5). Furthermore, a comparison between optical and inertial motion capture shows that sagittal kinematics are comparable (Chapter 6). Initial validation of a magnetic measurement system shows results that are promising for application in virtual reality training of stroke survivors (Appendix A).

The approaches presented in this thesis are applicable to many different applications. Furthermore, because a minimal sensor set is required it reduces costs for such applications significantly. In the discussion topics for future research that could further drive development of approaches for minimal sensing of body movements were indicated.

References

[1] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018. (Cited on pages 8 and 101).

[2] Y. Adesida, E. Papi, and A. H. McGregor, "Exploring the role of wearable technology in sport kinematics and kinetics: A systematic review," *Sensors (Switzerland)*, vol. 19, no. 7, 2019. (Cited on page 35).

[3] D. W. Aha, "Lazy learning," *Artificial Intelligence Review*, vol. 11, pp. 7–10, 1997. [Online]. Available: http://portal.acm.org/citation.cfm?id=273530.273552 (Cited on pages 8, 15, and 35).

[4] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2009. (Cited on pages 55 and 102).

[5] J. Amar, "Trottoir dynamographique," *Comptes rendus hebdomadaires des seances de l Academie des Sciences*, vol. 163, pp. 130–133, 1916. (Cited on pages 3 and 5).

[6] "Amfitrack," http://www.amfitrack.com/, AMIFITRACK, accessed: 2019-07-23. (Cited on page 132).

[7] A. Ancillao, S. Tedesco, J. Barton, and B. O'flynn, "Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review," *Sensors (Switzerland)*, vol. 18, no. 8, 2018. (Cited on page 35).

[8] A. Ancillao, M. M. van der Krogt, A. I. Buizer, M. M. Witbreuk, P. Cappa, and J. Harlaar, "Analysis of gait patterns pre- and post- Single Event Multilevel Surgery in children with Cerebral Palsy by means of Offset-Wise Movement Analysis Profile and Linear Fit Method," *Human Movement Science*, vol. 55, no. June, pp. 145–155, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.humov.2017.08.005 (Cited on page 35).

[9] S. Andrews, I. Huerta, T. Komura, L. Sigal, and K. Mitchell, "Real-time physics-based motion capture with sparse sensors," in *Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016)*, ser. CVMP

2016. New York, NY, USA: ACM, 2016, pp. 5:1–5:10. [Online]. Available: http://doi.acm.org/10.1145/2998559.2998564 (Cited on page 104).

[10] M. Andriluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall, and B. Schiele, "Posetrack: A benchmark for human pose estimation and tracking," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. (Cited on page 103).

[11] A. Aristidou and J. Lasenby, "Real-time marker prediction and cor estimation in optical motion capture," *The Visual Computer*, vol. 29, no. 1, pp. 7–26, Jan 2013. [Online]. Available: https://doi.org/10.1007/s00371-011-0671-y (Cited on page 7).

[12] Aristotle, *Aristotle: parts of animals, movement of animals, progression of animals*. Harvard University Press, 1968, translated by Peck A. Harvard. (Cited on page 3).

[13] E. R. Bachmann, I. Duman, U. Y. Usta, R. B. McGhee, X. Yun, and M. Zyda, "Orientation tracking for humans and robots using inertial sensors," *Proceedings 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation. CIRA'99 (Cat. No.99EX375)*, pp. 187–194, 1999. (Cited on page 6).

[14] N. I. Badler, N. I. Badler, M. J. Hollick, M. J. Hollick, J. P. Granieri, and J. P. Granieri, "Real time control of a virtual human using minimal sensors," *Presence*, vol. 2, no. 1, pp. 82–86, 1993. (Cited on pages 7, 15, and 103).

[15] G. P. Bailey and R. Harle, "Assessment of foot kinematics during steady state running using a foot-mounted IMU," *Procedia Engineering*, vol. 72, pp. 32–37, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.proeng.2014.06.009 (Cited on page 55).

[16] G. P. Bailey and R. K. Harle, "Measuring Temporal Parameters of Gait with Foot Mounted IMUs in Steady State Running," in *Proceedings of the 3rd International Congress on Sport Sciences Research and Technology Support - Volume 1: icSPORTS*, INSTICC. SciTePress, 2015, pp. 24–33. [Online]. Available: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005656500240033 (Cited on page 55).

[17] R. Baker, "The history of gait analysis before the advent of modern computers," *Gait & Posture*, vol. 26, no. 3, pp. 331 – 342, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636206003225 (Cited on page 3).

[18] M. Benedetti, A. Merlo, and A. Leardini, "Inter-laboratory consistency of gait analysis measurements," *Gait & Posture*, vol. 38, no. 4, pp. 934 – 939, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636213002130 (Cited on pages 89 and 96).

[19] J. H. M. Bergmann and A. H. McGregor, "Body-worn sensor design: What do patients and clinicians want?" *Annals of Biomedical Engineering*,

vol. 39, no. 9, pp. 2299–2312, Sep 2011. [Online]. Available: https://doi.org/10.1007/s10439-011-0339-9 (Cited on page 75).

[20] T. F. Besier, D. L. Sturnieks, J. A. Alderson, and D. G. Lloyd, "Repeatability of gait data using a functional hip joint centre and a mean helical knee axis," *Journal of Biomechanics*, vol. 36, no. 8, pp. 1159 – 1168, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929003000873 (Cited on page 96).

[21] M. Bianchi, P. Salaris, and A. Bicchi, "Synergy-based hand pose sensing: Optimal glove design," *The International Journal of Robotics Research*, vol. 32, no. 4, pp. 407–424, 2013. [Online]. Available: https://doi.org/10.1177/0278364912474079 (Cited on page 104).

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer-Verlag New York, 2006. (Cited on pages 8 and 40).

[23] M. C. Bisi, R. Stagni, A. Caroselli, and A. Cappello, "Anatomical calibration for wearable motion capture systems: Video calibrated anatomical system technique," *Medical Engineering & Physics*, vol. 37, no. 8, pp. 813 – 819, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350453315001319 (Cited on page 96).

[24] J. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307–310, 1986. (Cited on pages 59 and 80).

[25] A. Booth, M. van der Krogt, A. Buizer, F. Steenbrink, and J. Harlaar, "The validity and usability of an eight marker model for avatar-based biofeedback gait training," *Clinical Biomechanics*, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268003319300191 (Cited on page 7).

[26] G. A. Borelli, "On the movement of animals," in *Springer Berlin Heidelberg*, 1989. (Cited on page 3).

[27] A. Borrego, J. Latorre, M. Alcañiz, and R. Llorens, "Comparison of oculus rift and htc vive: Feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation," *Games for Health Journal*, vol. 7, no. 3, pp. 151–156, 2018. [Online]. Available: https://doi.org/10.1089/g4h.2017.0114 (Cited on page 132).

[28] J. Bortz, "A New Mathematical Formulation for Strapdown Inertial Navigation," *IEEE Transactions on Aerospace Electronic Systems*, vol. 7, pp. 61–66, Jan. 1971. (Cited on page 76).

[29] W. Braune and .-. Fischer, O. (Otto), *The human gait*. Berlin ; New York : Springer-Verlag, 1987, translation of: Der Gang des Menschen. (Cited on pages 3 and 4).

[30] B. Breine, P. Malcolm, I. Van Caekenberghe, P. Fiers, E. C. Frederick, and D. De Clercq, "Initial foot contact and related kinematics affect impact loading rate in running," *Journal of Sports Sciences*, vol. 35, no. 15, pp. 1556–1564, 2017. [Online]. Available: https://doi.org/10.1080/02640414.2016.1225970 (Cited on pages 92 and 96).

[31] I. V. Caekenberghe, V. Segers, P. Willems, T. Gosseye, P. Aerts, and D. D. Clercq, "Mechanics of overground accelerated running vs. running on an accelerated treadmill," *Gait & Posture*, vol. 38, no. 1, pp. 125–131, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636212003955 (Cited on page 55).

[32] G. Carlet, "Essai expérimental sur la locomotion humaine: Etude de la marche." Ph.D. dissertation, Paris, 1872. (Cited on pages 3 and 4).

[33] "CMU: Carnegie-Mellon Mocap Database." http://mocap.cs.cmu.edu, Carnegie-Mellon University, accessed: 13-07-2019. (Cited on pages 8, 48, 49, and 102).

[34] P. R. Cavanagh and M. A. Lafortune, "Ground Reaction Forces in Distance Running*," *Journal of Biomechanics*, vol. 13, no. 5, pp. 397–406, 1980. (Cited on pages 55 and 59).

[35] A. Cereatti, T. Bonci, M. Akbarshahi, K. Aminian, A. Barré, M. Begon, D. L. Benoit, C. Charbonnier, F. D. Maso, S. Fantozzi, C.-C. Lin, T.-W. Lu, M. G. Pandy, R. Stagni, A. J. van den Bogert, and V. Camomilla, "Standardization proposal of soft tissue artefact description for data sharing in human motion measurements," *Journal of Biomechanics*, vol. 62, pp. 5 – 13, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929017301008 (Cited on pages 89 and 96).

[36] J. Chai and J. K. Hodgins, "Performance animation from low-dimensional control signals," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 686, 2005. (Cited on pages 8, 15, 16, 21, 31, 35, 41, 103, and 132).

[37] J. H. Challis, "The variability in running gait caused by force plate targeting," *Journal of Applied Biomechanics*, vol. 17, no. 1, pp. 77–83, 2001. (Cited on page 55).

[38] S. Chanpimol, B. Seamon, H. Hernandez, M. Harris-Love, and M. R. Blackman, "Using xbox kinect motion capture technology to improve clinical rehabilitation outcomes for balance and cardiovascular health in an individual with chronic tbi," *Archives of Physiotherapy*, vol. 7, no. 1, p. 6, May 2017. [Online]. Available: https://doi.org/10.1186/s40945-017-0033-9 (Cited on page 35).

[39] E. Charry, W. Hu, M. Umer, A. Ronchi, and S. Taylor, "Study on estimation of peak ground reaction forces using tibial accelerations in running," in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2013, pp. 288–293. [Online]. Available: http:

//ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6529804 (Cited on pages 55, 58, 70, and 104).

[40] K. Chen, S. Gong, and T. Xiang, "Human pose estimation using structural support vector machines," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 846–851, 2011. (Cited on page 16).

[41] X. Chen and J. Davis, "Camera placement considering occlusion for robust motion capture," *Computer Graphics Laboratory, Stanford University, Tech. Rep*, vol. 2, 01 2002. (Cited on page 89).

[42] L. Chiari, U. D. Croce, A. Leardini, and A. Cappozzo, "Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors," *Gait & Posture*, vol. 21, no. 2, pp. 197 – 211, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636204000682 (Cited on page 5).

[43] K. P. Clark, L. J. Ryan, and P. G. Weyand, "Foot speed, foot-strike and footwear: linking gait mechanics and running ground reaction forces," *Journal of Experimental Biology*, vol. 217, no. 12, pp. 2037–2040, 2014. [Online]. Available: http://jeb.biologists.org/cgi/doi/10.1242/jeb.099523 (Cited on page 55).

[44] J. Cole, S. Crowle, G. Austwick, and D. H. Slater, "Exploratory findings with virtual reality for phantom limb pain; from stump motion to agency and analgesia," *Disability and Rehabilitation*, vol. 31, no. 10, pp. 846–854, 2009. [Online]. Available: https://doi.org/10.1080/09638280802355197 (Cited on page 35).

[45] D. M. Corey, W. P. Dunlap, and M. J. Burke, "Averaging Correlations: Expected Values and Bias in Combined Pearson *r*s and Fisher's *z* Transformations," *The Journal of General Psychology*, vol. 125, no. 3, pp. 245–261, 1998. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00221309809595548 (Cited on pages 59 and 92).

[46] H. P. Crowell and I. S. Davis, "Gait retraining to reduce lower extremity loading in runners," *Clinical Biomechanics*, vol. 26, no. 1, pp. 78–83, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.clinbiomech.2010.09.003 (Cited on pages 55, 59, and 60).

[47] M. Damsgaard, J. Rasmussen, S. T. Christensen, E. Surma, and M. de Zee, "Analysis of musculoskeletal systems in the anybody modeling system," *Simulation Modelling Practice and Theory*, vol. 14, no. 8, pp. 1100 – 1111, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1569190X06000554 (Cited on page 89).

[48] P. Davidson, H. Virekunnas, D. Sharma, R. Piché, and N. Cronin, "Continuous analysis of running mechanics by means of an integrated INS/GPS device," *Sensors (Switzerland)*, vol. 19, no. 6, pp. 1–20, 2019. (Cited on page 102).

[49] R. B. Davis, S. Õunpuu, D. Tyburski, and J. R. Gage, "A gait analysis data collection and reduction technique," *Human Movement Science*, vol. 10, no. 5, pp. 575 – 587, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016794579190046Z (Cited on pages 4, 35, 57, 89, and 91).

[50] W. de Vries, H. Veeger, C. Baten, and F. van der Helm, "Magnetic distortion in motion labs, implications for validating inertial magnetic sensors," *Gait & Posture*, vol. 29, no. 4, pp. 535 – 541, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636208003858 (Cited on pages 7, 75, and 89).

[51] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "OpenSim : Open-Source Software to Create and Analyze Dynamic Simulations of Movement," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 11, pp. 1940–1950, 2007. (Cited on pages 69, 89, and 91).

[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009. (Cited on page 103).

[53] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014. [Online]. Available: http://dx.doi.org/10.1561/2000000039 (Cited on page 9).

[54] P. Devita and W. a. Skelly, "Effect of landing stiffness on joint kinetics and energetics in the lower extremity." *Medicine and science in sports and exercise*, vol. 24, no. 1, pp. 108–115, 1992. (Cited on pages 55 and 59).

[55] J. P. A. Dewald, V. Sheshadri, M. L. Dawson, and R. F. Beer, "Upper-limb discoordination in hemiparetic stroke: Implications for neurorehabilitation," *Topics in Stroke Rehabilitation*, vol. 8, no. 1, pp. 1–12, 2001. [Online]. Available: https://doi.org/10.1310/WA7K-NGDF-NHKK-JAGD (Cited on page 135).

[56] V. Dietz, "Do human bipeds use quadrupedal coordination?" *Trends in Neurosciences*, vol. 25, no. 9, pp. 462–467, 2002. (Cited on page 19).

[57] D. Dinu, M. Fayolas, M. Jacquet, E. Leguy, J. Slavinski, and N. Houel, "Accuracy of postural human-motion tracking using miniature inertial sensors," *Procedia Engineering*, vol. 147, pp. 655 – 658, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705816307135 (Cited on page 89).

[58] D. D. D'Lima, B. J. Fregly, S. Patil, N. Steklov, and C. W. J. Colwell, "Knee joint forces: prediction, measurement, and significance," *Proceedings of the Institution of Mechanical Engineers, Part H*, vol. 226, no. 2, pp. 95–102, Feb 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22468461 (Cited on page 75).

[59] B. H. Dobkin, "Rehabilitation after stroke," *New England Journal of Medicine*, vol. 352, no. 16, pp. 1677–1684, 2005. [Online]. Available: https://doi.org/10.1056/NEJMcp043511 (Cited on page 132).

[60] T. W. Dorn, A. G. Schache, and M. G. Pandy, "Muscular strategy shift in human running: dependence of running speed on hip and ankle muscle performance," *Journal of Experimental Biology*, vol. 215, no. 13, pp. 2347–2347, 2012. [Online]. Available: https://jeb.biologists.org/content/215/13/2347 (Cited on page 91).

[61] E. Dorschky, M. Nitschke, A.-K. Seifer, A. J. van den Bogert, and B. M. Eskofier, "Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models," *Journal of Biomechanics*, p. 109278, aug 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0021929019304841?via{%}3Dihub (Cited on page 104).

[62] P. W. Duncan, K. J. Sullivan, A. L. Behrman, S. P. Azen, S. S. Wu, S. E. Nadeau, B. H. Dobkin, D. K. Rose, J. K. Tilson, S. Cen, and S. K. Hayden, "Body-weight–supported treadmill rehabilitation after stroke," *New England Journal of Medicine*, vol. 364, no. 21, pp. 2026–2036, 2011. [Online]. Available: https://doi.org/10.1056/NEJMoa1010790 (Cited on page 132).

[63] W. B. Edwards, T. R. Derrick, and J. Hamill, "Musculoskeletal attenuation of impact shock in response to knee angle manipulation," *Journal of Applied Biomechanics*, vol. 28, no. 5, pp. 502–510, 2012. (Cited on pages 55 and 59).

[64] P. Eichelberger, M. Ferraro, U. Minder, T. Denton, A. Blasimann, F. Krause, and H. Baur, "Analysis of accuracy in optical motion capture – a protocol for laboratory setup evaluation," *Journal of Biomechanics*, vol. 49, no. 10, pp. 2085 – 2088, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929016305681 (Cited on pages 3 and 104).

[65] G. S. Faber, C. C. Chang, I. Kingma, J. T. Dennerlein, and J. H. van Dieën, "Estimating 3D L5/S1 moments and ground reaction forces during trunk bending using a full-body ambulatory inertial motion capture system," *Journal of Biomechanics*, vol. 49, no. 6, pp. 904–912, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jbiomech.2015.11.042 (Cited on pages 55, 58, 59, and 70).

[66] W. Fang, L. Zheng, and J. Xu, "Self-contained optical-inertial motion capturing for assembly planning in digital factory," *The International Journal of Advanced Manufacturing Technology*, vol. 93, no. 1, pp. 1243–1256, Oct 2017. [Online]. Available: https://doi.org/10.1007/s00170-017-0526-4 (Cited on page 49).

[67] A. Ferrari, M. G. Benedetti, E. Pavan, C. Frigo, D. Bettinelli, M. Rabuffetti, P. Crenna, and A. Leardini, "Quantitative comparison of five current protocols in gait analysis," *Gait & Posture*, vol. 28, no. 2, pp. 207 – 216, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636207002743 (Cited on pages 36, 48, and 89).

[68] A. Ferrari, A. G. Cutti, P. Garofalo, M. Raggi, M. Heijboer, A. Cappello, and A. Davalli, "First in vivo assessment of "outwalk": a novel protocol for clinical gait analysis based on inertial and magnetic sensors," *Medical & Biological Engineering & Computing*, vol. 48, no. 1, p. 1, Nov 2009. [Online]. Available: https://doi.org/10.1007/s11517-009-0544-y (Cited on page 89).

[69] M. Ferrarin, G. Bovi, M. Rabuffetti, P. Mazzoleni, A. Montesano, E. Pagliano, A. Marchi, A. Magro, C. Marchesi, D. Pareyson, and I. Moroni, "Gait pattern classification in children with charcot-marie-tooth disease type 1a," *Gait & Posture*, vol. 35, no. 1, pp. 131–137, 2012. (Cited on page 96).

[70] M. Field, Z. Pan, D. Stirling, and F. Naghdy, "Human motion capture sensors and analysis in robotics," *Industrial Robot: An International Journal*, vol. 38, no. 2, pp. 163–171, 2011. (Cited on pages 15 and 69).

[71] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985. [Online]. Available: http://www.jneurosci.org/content/5/7/1688 (Cited on page 41).

[72] J. P. Folland, S. J. Allen, M. I. Black, J. C. Handsaker, and S. E. Forrester, "Running technique is an important component of running economy and performance," *Medicine and Science in Sports and Exercise*, vol. 49, no. 7, pp. 1412–1423, 2017. (Cited on page 55).

[73] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 4346–4354. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.494 (Cited on pages 36 and 41).

[74] M. Giuberti, "Inertial Sensing for Human Motion Analysis: Processing, Technologies, and Applications," Ph.D. dissertation, Università degli Studi di Parma, 2014. (Cited on pages 16, 20, and 21).

[75] J. F. Golding, "Motion sickness susceptibility," *Autonomic Neuroscience*, vol. 129, no. 1, pp. 67 – 76, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566070206002128 (Cited on pages 35 and 41).

[76] D. L. Goss, M. Lewek, B. Yu, W. B. Ware, D. S. Teyhen, and M. T. Gross, "Lower extremity biomechanics and self-reported foot-strike patterns among runners in traditional and minimalist shoes," *Journal of Athletic Training*, vol. 50, no. 6, pp. 603–611, 2012. (Cited on page 55).

[77] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649. (Cited on page 102).

[78] R. A. Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7297–7306. (Cited on page 103).

[79] S. R. Hamner, A. Seth, and S. L. Delp, "Muscle contributions to propulsion and support during running," *Journal of Biomechanics*, vol. 43, no. 14, pp. 2709 – 2716, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929010003611 (Cited on page 91).

[80] L. Herda, P. Fua, R. Plänkers, R. Boulic, and D. Thalmann, "Using skeleton-based tracking to increase the reliability of optical motion capture," *Human Movement Science*, vol. 20, no. 3, pp. 313 – 341, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167945701000501 (Cited on page 7).

[81] L. J. Hettinger and G. E. Riccio, "Visually induced motion sickness in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 3, pp. 306–310, 1992. [Online]. Available: https://doi.org/10.1162/pres.1992.1.3.306 (Cited on pages 35 and 41).

[82] M. Hirose, G. Deffaux, and Y. Nakagaki, "Development of an effective motion capture system based on data fusion and minimal use of sensors," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '96. New York, NY, USA: ACM, 1996, pp. 117–123. [Online]. Available: http://doi.acm.org/10.1145/3304181.3304204 (Cited on page 7).

[83] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. (Cited on pages 8, 39, and 102).

[84] R. G. Hoptroff, "The principles and practice of time series forecasting and business modelling using neural nets," *Neural Computing & Applications*, vol. 1, no. 1, pp. 59–66, Mar 1993. [Online]. Available: https://doi.org/10.1007/BF01411375 (Cited on page 48).

[85] A. Hreljac, "Impact and Overuse Injuries in Runners," *Medicine & Science in Sports & Exercise*, vol. 36, no. 5, pp. 845–849, 2004. (Cited on page 55).

[86] "Vive™ — discover virtual reality beyond imagination," https://www.vive.com/eu/, HTC, accessed: 2019-07-23. (Cited on page 135).

[87] Y. Huang, W. Jirattigalachote, M. R. Cutkosky, X. Zhu, and P. B. Shull, "Novel foot progression angle algorithm estimation via foot-worn, magneto-inertial sensing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2278–2285, Nov 2016. (Cited on pages 75, 83, and 102).

[88] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, vol. 37, pp. 185:1–185:15, Nov. 2018. (Cited on pages 8, 36, 37, 39, 40, 47, 48, 102, 103, and 104).

[89] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality*. Princeton, USA: Princeton University Press, 1999. (Cited on pages 20 and 38).

[90] D. Jack, R. Boian, A. S. Merians, M. Tremaine, G. C. Burdea, S. V. Adamovich, M. Recce, and H. Poizner, "Virtual reality-enhanced stroke rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 3, pp. 308–318, Sep. 2001. (Cited on page 132).

[91] J. Jerald, F. Brooks, Jr, M. C Advisor, B. D Whitton, A. , S. R Reader, G. Ellis, C. Bishop, A. A Member, C. Lastra, and M. , "Scene-motion-and latency-perception thresholds for head-mounted displays," Ph.D. dissertation, University of North Carolina, 05 2019. (Cited on pages 41 and 49).

[92] A. R. Jiménez, F. Seco, J. C. Prieto, and J. Guevara, "Indoor pedestrian navigation using an ins/ekf framework for yaw drift reduction and a foot-mounted imu," in *2010 7th Workshop on Positioning, Navigation and Communication*, March 2010, pp. 135–143. (Cited on pages 76 and 77).

[93] I. Jolliffe, *Principal Component Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1094–1096. (Cited on page 77).

[94] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten, "Measurement of lower extremity kinematics during level walking," *Journal of Orthopaedic Research*, vol. 8, no. 3, pp. 383–392, 1990. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jor.1100080310 (Cited on pages 57, 89, and 91).

[95] H. Kainz, L. Modenese, D. G. Lloyd, S. Maine, H. P. Walsh, and C. P. Carty, "Joint kinematic calculation based on clinical direct kinematic versus inverse kinematic gait models," *Journal of Biomechanics*, vol. 49, no. 9, pp. 1658–1669, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jbiomech.2016.03.052 (Cited on page 69).

[96] Y. Kangkang and P. Dinesh K, "FootSee : an Interactive Animation System," *2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pp. 329–339, 2003. (Cited on pages 16 and 103).

[97] A. Karatsidis, G. Bellusci, H. M. Schepers, M. de Zee, M. S. Andersen, and P. H. Veltink, "Estimation of ground reaction forces and moments during gait using only inertial motion capture," *Sensors*, vol. 17, no. 1, 2017. [Online]. Available: http://www.mdpi.com/1424-8220/17/1/75 (Cited on pages 35, 55, 58, 59, and 70).

[98] A. Karatsidis, M. Jung, H. M. Schepers, G. Bellusci, M. de Zee, P. H. Veltink, and M. S. Andersen, "Musculoskeletal model-based inverse dynamic analysis under ambulatory conditions using inertial motion capture," *Medical Engineering & Physics*, vol. 65, pp. 68 – 77, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350453319300189 (Cited on page 89).

[99] A. Karatsidis, R. E. Richards, J. M. Konrath, J. C. van den Noort, H. M. Schepers, G. Bellusci, J. Harlaar, and P. H. Veltink, "Validation of wearable visual feedback for retraining foot progression angle using inertial sensors and an augmented reality headset," *Journal of NeuroEngineering*

and Rehabilitation, vol. 15, no. 1, p. 78, Aug 2018. [Online]. Available: https://doi.org/10.1186/s12984-018-0419-2 (Cited on pages 35, 49, 75, 83, and 105).

[100] M. Khademi, H. Mousavi Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer, "Free-hand interaction with leap motion controller for stroke rehabilitation," in *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '14. New York, NY, USA: ACM, 2014, pp. 1663–1668. [Online]. Available: http://doi.acm.org/10.1145/2559206.2581203 (Cited on page 132).

[101] J. Kim, Y. Seol, and J. Lee, "Realtime Performance Animation Using Sparse 3D Motion Sensors," *Motion in Games*, pp. 31–42, 2012. (Cited on pages 16 and 103).

[102] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980 (Cited on page 40).

[103] E. Knippenberg, J. Verbrugghe, I. Lamers, S. Palmaers, A. Timmermans, and A. Spooren, "Markerless motion capture systems as training device in neurological rehabilitation: a systematic review of their use, application, target population and efficacy," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 61, Jun 2017. [Online]. Available: https://doi.org/10.1186/s12984-017-0270-x (Cited on page 35).

[104] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *International Joint Conference on Artificial Intelligence*, 1995. (Cited on pages 21 and 40).

[105] M. Kok, J. D. Hol, and T. B. Schön, "An optimization-based approach to human body motion capture using inertial sensors," *19th IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 79 – 85, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S147466701641596X (Cited on page 89).

[106] H. G. Kortier, J. Antonsson, H. M. Schepers, F. Gustafsson, and P. H. Veltink, "Hand pose estimation by fusion of inertial and magnetic sensing aided by a permanent magnet," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 796–806, Sep. 2015. (Cited on pages 104 and 132).

[107] B. Krüger, J. Baumann, M. Abdallah, and A. Weber, "A Study On Perceptual Similarity of Human Motions," *Workshop in Virtual Reality Interactions and Physical Simulation*, pp. 65–72, Dec 2011. (Cited on pages 21 and 41).

[108] B. Krüger, J. Tautges, A. Weber, and A. Zinke, "Fast Local and Global Similarity Searches in Large Motion Capture Databases," *Eurographics ACM SIGGRAPH Symposium on Computer Animation*, pp. 1–10, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?id=1921427.1921429 (Cited on page 16).

[109] J. P. Kulmala, J. Avela, K. Pasanen, and J. Parkkari, "Forefoot strikers exhibit lower running-induced knee loading than rearfoot strikers," *Medicine and Science in Sports and Exercise*, vol. 45, no. 12, pp. 2306–2313, 2013. (Cited on page 91).

[110] K. Kunze and P. Lukowicz, "Sensor placement variations in wearable activity recognition," *IEEE Pervasive Computing*, vol. 13, no. 4, pp. 32–41, Oct 2014. (Cited on page 104).

[111] I. Kutzner, A. Trepczynski, M. O. Heller, and G. Bergmann, "Knee adduction moment and medial contact force – facts about their correlation during gait," *PLOS ONE*, vol. 8, no. 12, pp. 1–8, 12 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0081036 (Cited on page 75).

[112] H. Kyröläinen, A. Belli, and P. Komi, "Biomechanical factors affecting running economy," *Medicine and Science in Sports and Exercise*, vol. 33, no. 8, pp. 1330–1337, 2001. (Cited on page 55).

[113] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M. J. Black, and P. V. Gehler, "Unite the people: Closing the loop between 3d and 2d human representations," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017. [Online]. Available: http://up.is.tuebingen.mpg.de (Cited on page 103).

[114] R. L. Lathrop, A. M. W. Chaudhari, and R. A. Siston, "Comparative assessment of bone pose estimation using point cluster technique and opensim," *Journal of Biomechanical Engineering*, vol. 133, pp. 1 – 2, 2011. (Cited on pages 89 and 96).

[115] K. Laver, B. Lange, S. George, J. Deutsch, G. Saposnik, and M. Crotty, "Virtual reality for stroke rehabilitation," *Cochrane Database of Systematic Reviews*, no. 11, 2017. (Cited on pages 105 and 132).

[116] A. Leardini, L. Chiari, U. D. Croce, and A. Cappozzo, "Human movement analysis using stereophotogrammetry: Part 3. soft tissue artifact assessment and compensation," *Gait & Posture*, vol. 21, no. 2, pp. 212 – 225, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636204000773 (Cited on page 89).

[117] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature14539 (Cited on pages 9 and 102).

[118] L. Legg, P. Langhorne, and O. S. Trialists, "Rehabilitation therapy services for stroke patients living at home: systematic review of randomised trials," *The Lancet*, vol. 363, no. 9406, pp. 352 – 356, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673604154342 (Cited on page 132).

[119] Z. Lin, Y. Xiong, H. Dai, and X. Xia, "An experimental performance evaluation of the orientation accuracy of four nine-axis mems motion sensors," in

2017 5th International Conference on Enterprise Systems (ES), Sep. 2017, pp. 185–189. (Cited on pages 47, 104, and 133).

[120] H. Liu, X. Wei, J. Chai, I. Ha, and T. Rhee, "Realtime human motion control with a small number of inertial sensors," *Symposium on Interactive 3D Graphics and Games*, pp. 133–140, 2011. (Cited on pages 16, 21, 31, and 103).

[121] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein, "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud," *Proc. VLDB Endow.*, vol. 5, no. 8, pp. 716–727, Apr. 2012. [Online]. Available: https://doi.org/10.14778/2212351.2212354 (Cited on page 49).

[122] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of Motion Capture as Surface Shapes," *arXiv:1904.03278*, 2019. (Cited on pages 8, 48, 49, and 102).

[123] E.-J. Marey, "De la mesure dans les differents acts de la locomotion.." Ph.D. dissertation, Comptes Rendues de l'Academie des Sciences de Paris, 1883. (Cited on pages 3 and 4).

[124] J. L. McGinley, R. Baker, R. Wolfe, and M. E. Morris, "The reliability of three-dimensional kinematic gait measurements: A systematic review," *Gait & Posture*, vol. 29, no. 3, pp. 360 – 369, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636208002646 (Cited on pages 4 and 89).

[125] D. McGrath, B. R. Greene, K. J. O'Donovan, and B. Caulfield, "Gyroscope-based assessment of temporal gait parameters during treadmill walking and running," *Sports Engineering*, vol. 15, no. 4, pp. 207–213, 2012. (Cited on page 55).

[126] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of vicon system positioning performance," *Sensors*, vol. 17, no. 7, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/7/1591 (Cited on page 89).

[127] C. E. Milner, J. Hamill, and I. Davis, "Are knee mechanics during early stance related to tibial stress fracture in runners?" *Clinical Biomechanics*, vol. 22, no. 6, pp. 697–703, 2007. (Cited on page 55).

[128] C. E. Milner and M. Paquette, "A kinematic method to detect foot contact during running for all foot strike patterns," *Journal of Biomechanics*, vol. 48, no. 12, pp. 1–4, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.jbiomech.2015.07.036 (Cited on page 58).

[129] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation Mocap Database HDM05," Universität Bonn, Tech. Rep. CG-2007-2, June 2007. (Cited on pages 8, 48, 49, and 102).

[130] F. Multon, L. France, M.-P. Cani-Gascuel, and G. Debunne, "Computer animation of human walking: a survey," *The Journal of Visualization and Computer Animation*, vol. 10, no. 1, pp. 39–54, 1999. (Cited on page 15).

[131] A. Mykhaylo, P. Leonid, r. G. Pete, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. (Cited on page 103).

[132] N. Nedergaard, M. Robinson, B. Drust, P. Lisboa, and J. Vanrenterghem, "Predicting ground reaction forces from trunk kinematics: A mass-spring-damper model approach," in *International Society of Biomechanics Conference Proceedings*, vol. 1, 2017, pp. 432–435. (Cited on pages 56 and 58).

[133] N. J. Nedergaard, J. Verheul, B. Drust, T. Etchells, P. Lisboa, M. A. Robinson, and J. Vanrenterghem, "The feasibility of predicting ground reaction forces during running from a trunk accelerometry driven mass-spring-damper model," *PeerJ*, vol. 6, p. e6105, Dec. 2018. [Online]. Available: https://doi.org/10.7717/peerj.6105 (Cited on page 104).

[134] J. Neter, M. Kutner, C. Nachtsheim, and W. Wasserman, *Applied Linear Statistical Models*. New-York: WCB McGraw-Hill, 1996. (Cited on page 80).

[135] B. M. Nigg, R. W. De Boer, and V. Fisher, "A kinematic comparison of overground and treadmill running," *Medicine and Science in Sports and Exercise*, vol. 27, pp. 98–105, 1995. (Cited on page 55).

[136] T. Novacheck, "The biomechanics of running." *Gait & posture*, vol. 7, no. 1, pp. 77–95, jan 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10200378 (Cited on pages 55 and 96).

[137] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (smij): A new representation for human skeletal action recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 24 – 38, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320313000680 (Cited on page 104).

[138] A. Oulasvirta, T. Roos, A. Modig, and L. Leppänen, "Information capacity of full-body movements," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 1289–1298. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466169 (Cited on page 104).

[139] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010. (Cited on page 103).

[140] G. Pavei, E. Seminati, J. L. Storniolo, and L. A. Peyré-Tartaruga, "Estimates of running ground reaction force parameters from motion analysis," *Journal of Applied Biomechanics*, vol. 33, no. 1, pp. 69–75, 2017. (Cited on pages 55, 57, 58, and 70).

[141] H. S. J. Picavet and J. M. W. Hazes, "Prevalence of self reported musculoskeletal diseases is high," *Annals of the Rheumatic Diseases*, vol. 62, no. 7, pp. 644–650, 2003. (Cited on page 75).

[142] C. Pizzolato, M. Reggiani, D. J. Saxby, E. Ceseracciu, L. Modenese, and D. G. Lloyd, "Biofeedback for Gait Retraining Based on Real-Time Estimation of Tibiofemoral Joint Contact Forces." *IEEE Transactions On Neural Systems And Rehabilitation Engineering: A Publication Of The IEEE Engineering In Medicine And Biology Society*, vol. 25, no. 9, pp. 1612–1621, 2017. (Cited on page 91).

[143] G. B. Prange, M. J. Jannink, C. G. Groothuis-Oudshoorn, H. J. Hermens, and M. J. IJzerman, "Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke," *Journal of Rehabilitation Research and Development*, vol. 43, pp. 171–183, 2006. (Cited on page 132).

[144] G. B. Prange-Lasonder, B. Radder, A. I. R. Kottink, A. Melendez-Calderon, J. H. Buurke, and J. S. Rietman, "Applying a soft-robotic glove as assistive device and training tool with games to support hand function after stroke: Preliminary results on feasibility and potential clinical impact," in *2017 International Conference on Rehabilitation Robotics (ICORR)*, July 2017, pp. 1401–1406. (Cited on page 135).

[145] "Pti phoenix technologies 3d motion capture systems homepage," http://www.ptiphoenix.com/, PTI Phoenix Technologies, accessed: 2016-05-01. (Cited on page 15).

[146] "Qualisys homepage," http://www.qualisys.com/, Qualisys, accessed: 2019-04-01. (Cited on pages 15, 35, 89, and 132).

[147] E. Ragan, C. Wilkes, D. A. Bowman, and T. Hollerer, "Simulation of augmented reality systems in purely virtual environments," in *2009 IEEE Virtual Reality Conference*, March 2009, pp. 287–288. (Cited on page 138).

[148] J. R. Rambach, A. Tewari, A. Pagani, and D. Stricker, "Learning to fuse: A deep learning approach to visual-inertial camera pose estimation," in *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Sep. 2016, pp. 71–76. (Cited on page 36).

[149] J. Reenalda, E. Maartens, L. Homan, and J. J. Buurke, "Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics," *Journal of Biomechanics*, vol. 49, no. 12, 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S002192901630968X (Cited on pages 7, 55, 69, 90, 91, 92, and 102).

[150] N. D. Reeves and F. L. Bowling, "Conservative biomechanical strategies for knee osteoarthritis," *Nature Reviews Rheumatology*, vol. 7, pp. 113–122, Feb 2011. [Online]. Available: https://doi.org/10.1038/nrrheum.2010.212 (Cited on page 75).

[151] L. Ren, R. K. Jones, and D. Howard, "Whole body inverse dynamics over a complete gait cycle based only on measured kinematics," *Journal of Biomechanics*, vol. 41, no. 12, pp. 2750–2759, 2008. (Cited on page 59).

[152] Q. Riaz, G. Tao, B. Krüger, and A. Weber, "Motion reconstruction using very few accelerometers and ground contacts," *Graphical Models*, vol. 79, pp. 23–38, 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1524070315000107 (Cited on pages 15, 31, and 103).

[153] R. Richards, M. van der Esch, J. C. van den Noort, and J. Harlaar, "The learning process of gait retraining using real-time feedback in patients with medial knee osteoarthritis," *Gait & Posture*, vol. 62, pp. 1 – 6, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S096663621830105X (Cited on page 83).

[154] R. E. Richards, J. C. van den Noort, M. van der Esch, M. J. Booij, and J. Harlaar, "Effect of real-time biofeedback on peak knee adduction moment in patients with medial knee osteoarthritis: Is direct feedback effective?" *Clinical Biomechanics*, vol. 57, pp. 150 – 158, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268003317301419 (Cited on page 75).

[155] P. O. Riley, J. Dicharry, J. Franz, U. D. Croce, R. P. Wilder, and D. C. Kerrigan, "A kinematics and kinetic comparison of overground and treadmill running," *Medicine and Science in Sports and Exercise*, vol. 40, no. 6, pp. 1093–1100, 2008. (Cited on pages 55 and 91).

[156] E. Ringdahl and S. Pandit, "Treatment of knee osteoarthritis," *Am Fam Physician*, vol. 83, no. 11, pp. 1287–1292, Jun 2011. (Cited on page 75).

[157] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon, "Accuracy and repeatability of single-pose calibration of inertial measurement units for whole-body motion analysis," *Gait & Posture*, vol. 54, pp. 80 – 86, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636217300723 (Cited on pages 89 and 96).

[158] D. G. E. Robertson, "Wikipedia - motion capture," https://commons.wikimedia.org/w/index.php?curid=6896127, accessed: 2019-08-30. (Cited on page 5).

[159] D. Roetenberg, P. Slycke, A. Ventevogel, and P. Veltink, "A portable magnetic position and orientation tracker," *Sensors and Actuators A: Physical*, vol. 135, no. 2, pp. 426 – 432, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092442470600553X (Cited on pages 104 and 132).

[160] D. Roetenberg, P. J. Slycke, and P. H. Veltink, "Ambulatory position and orientation tracking fusing magnetic and inertial sensing," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 883–890, May 2007. (Cited on pages 6, 104, and 132).

[161] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN : Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors," *Xsens Technologies*, pp. 1–9, 2013. (Cited on pages 6, 15, 16, 18, 55, 56, and 89).

[162] H. Rouhani, J. Favre, X. Crevoisier, and K. Aminian, "Ambulatory assessment of 3D ground reaction force using plantar pressure distribution," *Gait and Posture*, vol. 32, no. 3, pp. 311–316, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.gaitpost.2010.05.014 (Cited on page 69).

[163] D. Rutherford, C. Hubley-Kozey, K. Deluzio, W. Stanish, and M. Dunbar, "Foot progression angle and the knee adduction moment: a cross-sectional investigation in knee osteoarthritis," *Osteoarthritis and Cartilage*, vol. 16, no. 8, pp. 883 – 889, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063458407003779 (Cited on pages 75 and 76).

[164] A. M. Sabatini, "Quaternion-based strap-down integration method for applications of inertial sensing to gait analysis," *Medical and Biological Engineering and Computing*, vol. 43, no. 1, pp. 94–101, Feb 2005. [Online]. Available: https://doi.org/10.1007/BF02345128 (Cited on pages 6 and 75).

[165] A. Safonova, J. K. Hodgins, and N. S. Pollard, "Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces," *ACM Transactions on Graphics*, vol. 23, no. 3, p. 514, 2004. (Cited on pages 7, 15, 35, and 38).

[166] T. D. Sanger, "Human arm movements described by a low-dimensional superposition of principal components." *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 20, no. 3, pp. 1066–1072, 2000. (Cited on pages 7, 15, 35, and 38).

[167] H. M. Schepers, H. F. J. M. Koopman, and P. H. Veltink, "Ambulatory assessment of ankle and foot dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 895–902, May 2007. (Cited on page 78).

[168] H. M. Schepers, H. J. Luinge, G. Bellusci, and P. Slycke, "XKF3 - Low-Power, Optimal Estimation of 3D Orientation using Inertial and Magnetic Sensing," *Xsens*, pp. 1–9, 2015. (Cited on pages 6 and 18).

[169] H. M. Schepers and P. H. Veltink, "Stochastic magnetic measurement model for relative position and orientation estimation," *Measurement Science and Technology*, vol. 21, no. 6, jun 2010. (Cited on pages 104, 132, and 139).

[170] M. Schepers, M. Giuberti, and G. Bellusci, "Xsens MVN : Consistent Tracking of Human Xsens MVN : Consistent Tracking of Human Motion Using Inertial Sensing," *Xsens*, 2018. (Cited on pages 5, 7, 35, 47, 89, 91, 96, and 104).

[171] A. Schmitz, M. B. Pohl, K. Woods, and B. Noehren, "Variables during swing associated with decreased impact peak and loading rate in running," *Journal of Biomechanics*, vol. 47, no. 1, pp. 32–38, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jbiomech.2013.10.026 (Cited on pages 55 and 59).

[172] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Processing*, vol. 45, pp. 2673–2681, 1997. (Cited on page 39).

[173] T. Seel, J. Raisch, and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors (Basel, Switzerland)*, vol. 14, no. 4, pp. 6891–6909, 2014. (Cited on page 18).

[174] S. K. Semwal, R. Hightower, and S. Stansfield, "Mapping algorithms for real-time control of an avatar using eight sensors," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 1, pp. 1–21, 1998. (Cited on page 15).

[175] S. Sharif Bidabadi, I. Murray, and G. Y. F. Lee, "Validation of foot pitch angle estimation using inertial measurement unit against marker-based optical 3D motion capture system," *Biomedical Engineering Letters*, vol. 8, no. 3, pp. 283–290, 2018. [Online]. Available: https://doi.org/10.1007/s13534-018-0072-5 (Cited on page 83).

[176] J. Shippen and B. May, "A kinematic approach to calculating ground reaction forces in dance." *Journal of dance medicine & science : official publication of the International Association for Dance Medicine & Science*, vol. 16, no. 1, pp. 39–43, 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22390953 (Cited on pages 55, 58, and 70).

[177] P. B. Shull, W. Jirattigalachote, M. A. Hunt, M. R. Cutkosky, and S. L. Delp, "Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait & Posture*, vol. 40, no. 1, pp. 11 – 19, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966636214002872 (Cited on page 75).

[178] P. B. Shull, R. Shultz, A. Silder, J. L. Dragoo, T. F. Besier, M. R. Cutkosky, and S. L. Delp, "Toe-in gait reduces the first peak knee adduction moment in patients with medial compartment knee osteoarthritis," *Journal of Biomechanics*, vol. 46, no. 1, pp. 122 – 128, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929012006173 (Cited on pages 75 and 80).

[179] M. Simic, T. Wrigley, R. Hinman, M. Hunt, and K. Bennell, "Altering foot progression angle in people with medial knee osteoarthritis: the effects of varying toe-in and toe-out angles are mediated by pain and malalignment," *Osteoarthritis and Cartilage*, vol. 21, no. 9, pp. 1272 – 1280, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063458413008376 (Cited on page 75).

[180] E. Simo-Serra, A. Ramisa, G. Alenyà, C. Torras, and F. Moreno-Noguer, "Single image 3D human pose estimation from noisy observations," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 2673–2680. (Cited on page 47).

[181] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. (Cited on page 102).

[182] J. Sinclair, J. I. M. Richards, P. J. Taylor, C. J. Edmundson, D. Brooks, and J. Sarah, "Three-dimensional kinematic comparison of treadmill and overground running," *Sports Biomechanics*, no. October 2014, pp. 37–41, 2013. (Cited on page 55).

[183] I. Skog, P. Handel, J. Nilsson, and J. Rantakokko, "Zero-velocity detection—an algorithm evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2657–2666, Nov 2010. (Cited on pages 7 and 76).

[184] L. H. Sloot, H. Houdijk, and J. Harlaar, "A comprehensive protocol to test instrumented treadmills," *Medical Engineering and Physics*, vol. 37, no. 6, pp. 610–616, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.medengphy.2015.03.018 (Cited on page 57).

[185] R. Slyper, J. K. Hodgins, R. Slyper, and J. Hodgins, "Action capture with accelerometers," *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 193–199, 2008. (Cited on pages 15, 16, 35, and 104).

[186] R. Stagni, S. Fantozzi, A. Cappello, and A. Leardini, "Quantification of soft tissue artefact in motion analysis by combining 3d fluoroscopy and stereophotogrammetry: a study on two subjects," *Clinical Biomechanics*, vol. 20, no. 3, pp. 320 – 329, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268003304002839 (Cited on page 96).

[187] "Steamvr," https://www.steamvr.com/en/, Steam, accessed: 2019-07-29. (Cited on page 134).

[188] F. Stief, H. Böhm, K. Michel, A. Schwirtz, and L. Döderlein, "Reliability and Accuracy in Three-Dimensional Gait Analysis : A Comparison of Two Lower Body Protocols," *Journal of applied biomechanics*, pp. 105–111, 2013. (Cited on pages 36, 48, 69, and 89).

[189] C. Strohrmann, H. Harms, G. Tröster, S. Hensler, and R. Müller, "Out of the lab and into the woods: Kinematic analysis in running using wearable sensors," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, ser. UbiComp '11. New York, NY, USA: ACM, 2011, pp. 119–122. [Online]. Available: http://doi.acm.org/10.1145/2030112.2030129 (Cited on page 91).

[190] Y. Suzuki, T. Inoue, and T. Nomura, "A simple algorithm for assimilating marker-based motion capture data during periodic human movement into models of multi-rigid-body systems," *Frontiers in Bioengineering and Biotechnology*, vol. 6, p. 141, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fbioe.2018.00141 (Cited on page 7).

[191] M. P. Tartaruga, J. Brisswalter, L. A. Peyré-Tartaruga, A. O. V. Ávila, C. L. Alberton, M. Coertjens, E. L. Cadore, C. L. Tiggemann, E. M. Silva, and L. F. M. Kruel, "The relationship between running economy and biomechanical variables in distance runners," *Research Quarterly for Exercise and Sport*, vol. 83, no. 3, pp. 367–375, 2012. (Cited on page 55).

[192] J. Tautges, A. Zinke, B. Krüger, J. Baumann, A. Weber, T. Helten, M. Müller, H.-P. Seidel, and B. Eberhardt, "Motion Reconstruction Using Sparse Accelerometer Data," *ACM Transactions on Graphics*, vol. 30, no. 3, pp. 18:1–12, 2011. (Cited on pages 8, 15, 16, 21, 24, 27, 31, 35, 37, 55, 104, and 132).

[193] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Computer Vision and Pattern Recognition*, 2014. (Cited on page 103).

[194] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 4489–4497. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.510 (Cited on page 36).

[195] N. F. Troje, "Decomposing biological motion: a framework for analysis and synthesis of human gait patterns." *Journal of vision*, vol. 2, no. 5, pp. 371–387, 2002. (Cited on pages 7, 15, 35, and 38).

[196] D. Tzovaras, N. Ploskas, and M. Strintizis, "Rigid 3-D motion estimation using neural networks and initially estimated 2-D motion data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 1, pp. 158–165, 2000. (Cited on page 16).

[197] "Unity real-time development platform — 3d, 2d vr & ar visualizations," https://unity.com/, UNITY, accessed: 2019-07-29. (Cited on page 134).

[198] R. N. van Gent, D. Siem, M. van Middelkoop, A. G. van Os, S. M. A. Bierma-Zeinstra, and B. W. Koes, "Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review," *British Journal of Sports Medicine*, vol. 41, no. 8, pp. 469–480, 2007. [Online]. Available: http://bjsm.bmj.com/content/41/8/469 (Cited on page 55).

[199] F. B. van Meulen, J. Reenalda, J. H. Buurke, and P. H. Veltink, "Assessment of daily-life reaching performance after stroke," *Annals of Biomedical Engineering*, vol. 43, no. 2, pp. 478–486, Feb 2015. [Online]. Available: https://doi.org/10.1007/s10439-014-1198-y (Cited on page 105).

[200] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, "Learning from synthetic humans," in *CVPR*, 2017. (Cited on page 103).

[201] "Vicon homepage," http://www.vicon.com/, Vicon, accessed: 2016-04-01. (Cited on pages 4, 5, 15, 35, 89, and 132).

[202] T. Von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll, "Sparse Inertial Poser : Automatic 3D Human Pose Estimation from Sparse IMUs," *Computer Graphics Forum*, vol. 36, no. 2, pp. 349–360, may 2017. (Cited on pages 7, 36, 37, 41, 102, 103, and 104).

[203] W. Weber and E. Weber, *Mechanics of the Human Walking Apparatus*. Berlin: Springer-Verlag, 1992, translated by Maquet, P and Furlong, R. (Cited on page 3).

[204] Wikipedia, "Wikipedia - artificial neural network," https://en.wikipedia.org/wiki/Artificial_neural_network, accessed: 2019-09-06. (Cited on page 9).

[205] R. Willy, M. B. Pohl, and I. S. Davis, "Calculation of vertical load rates in the absence of vertical impact peaks," in *American Society of Biomechanics Meeting*, 2008. (Cited on page 60).

[206] D. A. Winter, *Biomechanics and motor control of human movement*, 4th ed. New York, USA: John Wiley & Sons, 2009. (Cited on pages 21 and 104).

[207] A. D. Woolf and B. Pfleger, "Burden of major musculoskeletal conditions," *Bull. World Health Organ.*, vol. 81, no. 9, pp. 646–656, 2003. (Cited on page 75).

[208] F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. F. Van Beijnum, and P. H. Veltink, "On the Validity of Different Motion Capture Technologies for the Analysis of Running," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Aug 2018, pp. 1175–1180. (Cited on pages 5, 36, and 48).

[209] F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B.-J. F. van Beijnum, and P. H. Veltink, "Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors," *Frontiers in Physiology*, vol. 9, p. 218, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fphys.2018.00218 (Cited on pages 35, 37, and 49).

[210] F. J. Wouda, M. Giuberti, G. Bellusci, and P. H. Veltink, "Estimation of full-body poses using only five inertial sensors: An eager or lazy learning approach?" *Sensors*, vol. 16, no. 12, 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/12/2138 (Cited on pages 8, 35, 36, 37, 38, 41, 47, 48, 49, 55, 56, 59, 69, and 132).

[211] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D. D. D'Lima, L. Cristofolini, H. Witte, O. Schmid, and I. Stokes, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine," *Journal of Biomechanics*, vol. 35, no. 4, pp. 543–548, 2002. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0021929001002226 (Cited on page 57).

[212] H. Xia, J. Xu, J. Wang, M. A. Hunt, and P. B. Shull, "Validation of a smart shoe for estimating foot progression angle during walking gait," *Journal of Biomechanics*, vol. 61, pp. 193 – 198, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929017303718 (Cited on page 83).

[213] Y. Xiang, H.-J. Chung, J. H. Kim, R. Bhatt, S. Rahmatalla, J. Yang, T. Marler, J. S. Arora, and K. Abdel-Malek, "Predictive dynamics: an

optimization-based novel approach for human motion simulation," *Structural and Multidisciplinary Optimization*, vol. 41, no. 3, pp. 465–479, Apr 2010. [Online]. Available: https://doi.org/10.1007/s00158-009-0423-z (Cited on page 36).

[214] "Mvn - products - xsens 3d motion tracking," https://www.xsens.com/products/xsens-mvn/, Xsens, accessed: 2019-04-01. (Cited on pages 6, 15, 16, 18, 35, 89, and 132).

[215] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. (Cited on page 102).

[216] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833. (Cited on page 101).

[217] J.-t. Zhang, A. C. Novak, B. Brouwer, and Q. Li, "Concurrent validation of Xsens MVN measurement of lower limb joint angular kinematics." *Physiological measurement*, vol. 34, no. 8, pp. N63–9, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23893094 (Cited on page 18).

[218] H. Zheng, R. J. Davies, and N. D. Black, "Web-based monitoring system for home-based rehabilitation with stroke patients," in *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*, June 2005, pp. 419–424. (Cited on page 132).

[219] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei, "Towards 3d human pose estimation in the wild: A weakly-supervised approach," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 398–407. (Cited on page 102).

[220] Y. Zhuang, J. Gong, D. C. Kerrigan, B. C. Bennett, J. Lach, and S. Russell, "Gait tracker shoe for accurate step-by-step determination of gait parameters," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, June 2016, pp. 13–18. (Cited on page 83).

Appendices

Appendix

A

Magnetic Sensing For Estimation of End-Effector Position in a Virtual Reality Application

Abstract

Current rehabilitation of stroke survivors requires extensive support of physical therapists and/or clinical experts. Part of this rehabilitation could be performed in the patient's home when progress could be monitored sufficiently by their doctor. However, current eHealth applications do not provide doctors with sufficient qualitative information of the patient's progress. Therefore, we propose a VR application for training of reaching tasks using only a magnetic actuator and sensor placed at the chest and hand, respectively. This would allow patients to train in their home environment with a sensor system that is set up easily and fast. The following research question will be investigated: "Can magnetic sensing be used to accurately drive a virtual training environment for reaching tasks?". Measurements using the magnetic sensor setup show differences of approximately 3 cm in relative position estimation compared to an optical reference system. This demonstrates that the magnetic measurement technology has potential for this type of rehabilitation application. However, the developed mock-up VR application requires improved measurement of the heading direction to align the measurement system and the virtual environment.

Keywords: Virtual Reality; Magnetic Sensors; Inertial Sensors; Minimal Sensing; Stroke Rehabilitation.

A.1 Introduction

Stroke survivors typically require extensive physical and cognitive rehabilitation under supervision of physical therapists and/or specialists to improve their quality of life [59]. Physical rehabilitation is an important step for stroke survivors to improve their independence. However, it requires significant support from physical therapists. Therefore, supportive devices have been developed to reduce work load on physical therapists, e.g., weight-supported treadmills or robotic gait training devices [62, 143]. Additionally, effective physical therapy can be applied in the home environment [118]. This led to the development of e-health applications for stroke patients [218]. Currently, such applications require users to follow a home training regime, but are still expected to regularly visit their doctor to monitor their progress.

Virtual reality (VR) has shown to be an effective environment for training upper limb function [90, 115]. The use of virtual reality opens up unlimited opportunities for training scenarios, e.g., various environments could be rendered based on the patients' progress or the virtual and actual world could be mismatched to improve specific training outcomes. However, most motion capture systems require external infrastructure, e.g., HTC Vive [27], Qualisys [146]) and Vicon [201]), or significant computing power, e.g., LEAP motion [100]. Furthermore, inertial motion capture can be used in any environment, but requires a large number of body-worn sensors (e.g., Xsens MVN [214]). Efforts to reduce this number of markers/sensors have shown to be effective for generic applications [36, 192, 210]. However, specific modifications would be necessary to obtain the required accuracy for training applications for stroke survivors.

Magnetic actuation and sensing has the potential to overcome these limitations, as it provides relative position without requiring external infrastructure, a large number of sensors or significant computing power. This principle can be applied using either a powered [159, 160, 169] or a permanent magnet [106]. The benefit of using a powered magnet is that no or only limited additional information or assumptions are required to estimate the 3D position and orientation of a 3D magnetic sensor relative to the 3D actuator when time-division [159, 160] or frequency-division multiplexing is applied to the magnetic field in the different measurement directions. Also, multiple magnetic actuators can be applied if desired, when the magnetic field is properly actuated. However, this method requires a power source, which is also required for VR equipment.

To the best of our knowledge, there is no minimal sensing approach for measuring end-effector position in a virtual environment with only on-body sensing/actuation and processing. This leads to the following research question: "Can magnetic sensing be used to accurately drive a virtual training environment for reaching tasks?". We chose to use a magnetic sensor module of Amfitrack [6], due to its modular design (multiple sensors can be added) and implementation of a sufficient API. Accuracy of this system is unknown (not provided by the manufacturer), therefore, the following sub question will be answered: "How accurate is the relative position measurement of the magnetic sensing system?". Additionally, we will address the following goal: "Design of an effective virtual training environment for reaching tasks".

A.2 Methods

In the following two sections, the methods to assess the static and dynamic accuracy of the Amfitrack system (Section A.2.1) and the initial design of the VR environment to assess the feasibility of training reaching tasks for stroke survivors (Section A.2.2).

A.2.1 Accuracy Measurements

The Amfitrack system measures relative position and orientation by measuring the actuated magnetic field from the source. Because of perpendicular placement of the actuation axes within the source in combination with time-division multiplexing, three-dimensional relative (to the source) position and orientation can be obtained. The accuracy of such measurements are not qualitatively stated by the manufacturer, hence this is evaluated in this work.

Position accuracy of the Amfitrack system is analyzed by comparing estimated relative positions with those from using an optical reference system. To effectively determine this accuracy, locations of the actuator and sensor within the casing should be known to provide a fair accuracy measure. Furthermore, the Amfitrack system measures quantities relative to the source, hence the orientation of the source itself is unknown. This orientation is required to align the Amfitrack coordinate frame with the virtual environment.

Figure A.1 shows the measurement setup, with a magnetic actuator (**A**) and sensor (**S**). Relative to the actuator, position (in millimeters) and orientation (in quaternions) of the magnetic sensor was recorded at 125 Hz using Amfitrack Viewer (Amfitrack, Vejle, Denmark). Position (in millimeters) of the three retro-reflective markers on both the actuator and sensor was measured at 100 Hz by eight high-speed infrared cameras (Vicon, Oxford, UK) and processed using Nexus 2.8.2 (Vicon, Oxford, UK). This marker setup was chosen to obtain the locations of the actuator and sensor within the casing. An Inertial Measurement Unit (IMU, Xsens, Enschede, the Netherlands) was attached to the magnetic actuator to obtain its' orientation at 100 Hz, expressed in quaternions, which is accurate within 0.5 degrees [119]. With this orientation the Amfitrack coordinate frame can be aligned with the virtual environment. This can be determined based on the optical markers, however, in a VR application this optical information source will not be available. Therefore, this setup included an IMU to test if this would be an appropriate sensor for determining orientation of the actuator. To synchronize all three different measurement systems the magnetic actuator was rotated twice around its' vertical axis, this rotation is present in all signals. Furthermore, the Amfitrack data was downsampled to 100 Hz to match the sample frequency of the other measurement systems.

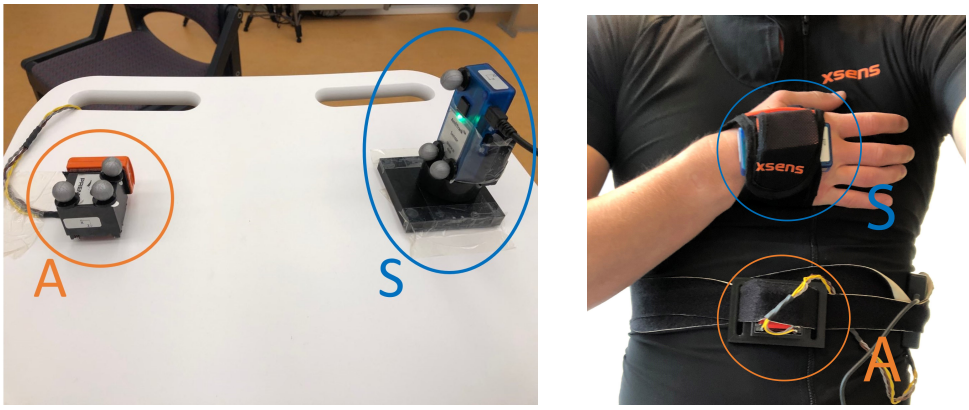
The following three experimental conditions were measured for evaluation of the position accuracy:

1. Static measurement, i.e., actuator/sensor are not moved w.r.t. each other. Magnetic disturbances (ferromagnetic material, scissors and a phone) are applied in between the actuator and sensor.

2. Movement of the magnetic sensor along all three axes individually (at a constant velocity) to a distance of approximately 2 meters.

3. Random reaching movements while wearing the system on-body, with exploration of the whole feasible workspace of the hand relative to the trunk.

This is evaluated using both the Euclidean distance and Root Mean Square Error (RMSE) over all three axes.



(a) Static measurement setup. (b) On-body measurement setup.

Figure A.1: The measurement setup used for validating the position accuracy of the Amfitrack system. (**A**) is the magnetic actuator that is powered by USB, and (**S**) is the magnetometer. In A.1(a) 3 optimal markers are placed on both the magnetic actuator and sensor for measuring their 3D positions and orientations with an optical reference system. In A.1(b) the actuator and sensor with retro-reflective markers were placed on the body using Velcro straps (markers were placed after the picture was taken). In this manner, the relative position of the sensor can be estimated w.r.t. a steady position on the body. In both cases an IMU is placed on the magnetic actuator, to determine if this information source can provide an accurate estimate of the actuator orientation, such that the obtained relative positions can be rotated to a global coordinate frame, which is required for a VR application.

A.2.2 VR Application Design

Based on discussions with a physical therapist expert, we came to the following requirements for a virtual training environment for upper-body movements of stroke survivors:

- Simple and clean environment, such that users can focus on the task at hand without unnecessary distracting details.
- Tasks should be challenging and relevant for upper-body functional rehabilitation of stroke survivors. E.g., reaching for a cup in the kitchen or catching a ball.
- Qualitative evaluation of the performed tasks should be implemented and accessible by the user's doctor.

An initial prototype of the virtual environment was built using UNITY [197] (to visualize the different elements in the environment and run the necessary calculations such as aligning the coordinate frames) in combination with SteamVR [187]

(to enable use of a VR headset). The initial prototype environment (shown in Figure A.2) displays a neutral environment with a table and blue ball. The goal of this rehabilitation application is to improve the kinematics of stroke survivors for reaching tasks, since this is hindered by pathological synergies [55]. Such an application could potentially be combined with passive arm-support [144]. A training environment benefits from a minimal sensor configuration as it is minimally obtrusive and can be donned and doffed more quickly by patients themselves.

In this initial prototype, a virtual hand is shown that represents position of the user's hand. Whenever the blue ball (which is 7 cm in diameter) is touched by the user, its' location changes such that a different reaching task can be performed. It might be beneficial to include a movement back to a neutral position, e.g., the user's chest, after each reaching task, such that these are more representative of tasks from daily life. Furthermore, size of the ball could be adaptable to the patients' progression, i.e., use a smaller ball when the patient has improved his/her movement control more. This would however pose higher accuracy requirements on the system, as a position error is more observable if the ball is of smaller size.

The user is shown (with numbers in the background) how much distance, time and with what velocity consecutive balls are touched. For demonstration purposes a video of this concept environment can be found in section A.6. This concept meets all three requirements, however, performance of this concept with either HTC Vive controller or magnetic measurement system inputs still needs to be evaluated. We chose to use the VR glasses of HTC VIVE [86], because of its' compatibility with UNITY and SteamVR.

Figure A.2: Concept application for training reaching tasks of stroke survivors. The goal is to touch the blue ball, which moves after a touch. The time, distance and velocity between different targets are displayed as feedback to the user. In this concept these are presented as numbers, however, a more visual representation (e.g., using bars or pie-charts) might be more effective. Furthermore, velocity is shown as infinity here because of the initialization, i.e., no ball has been touched yet.

A.3 Results

The resulting position measurements during a static trial (1, as defined in section A.2.1) of the Amfitrack measurement system and the optical reference system are shown in Figure A.3. The observed variations in the measured coordinates ($RMSE_x = 0.005$ m, $RMSE_y = 0.010$ m and $RMSE_z = 0.018$ m, between 8 and 25 seconds) with Amfitrack are the effect of disturbing the magnetic field of the actuator by moving scissors and a phone in between the sensor and actuator. This results in differences as large as 10 cm for the z-axis in this measurement. This variation is smaller if we look at the Euclidean distance ($RMSE_{euclidean} = 0.006$ m, between 8 and 25 seconds, in the right part of Figure A.3). However, such differences only occurred if the magnetic field is disturbed in between the actuator and sensor, since in the intermediate time periods the disturbance was approximately 25 cm away from the magnetic sensor. These results indicate that impact of magnetic disturbances is minimal unless these are in between the actuator and sensor.

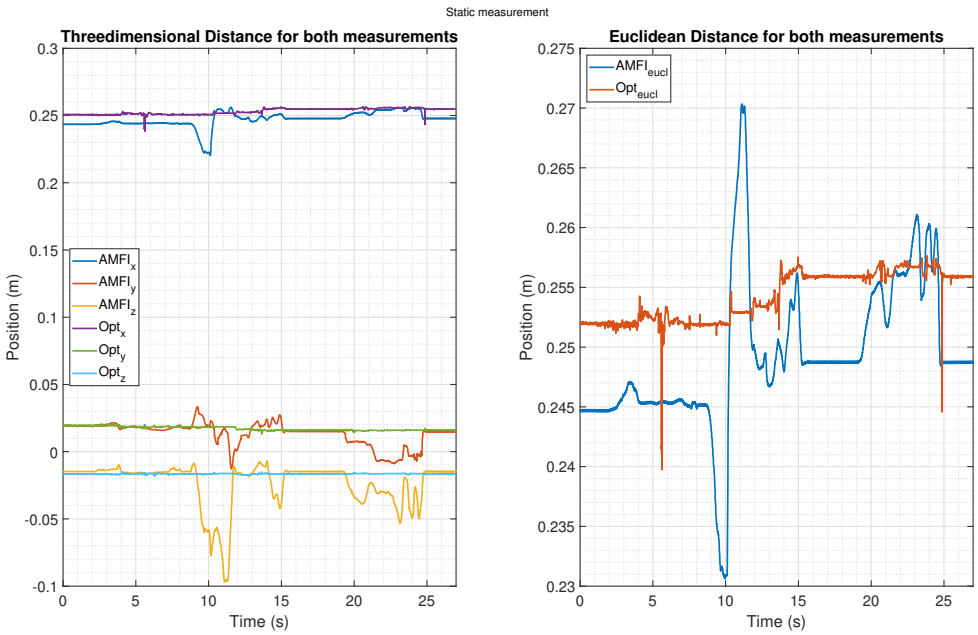


Figure A.3: During this trial both the actuator and sensor were static, but magnetic disturbances were applied in between the actuator and sensor (in the form of scissors and a phone). Position of the magnetic sensor relative to the actuator measured with both Amfitrack (AMFI) and Vicon (Opt). (**Left**) shows the comparison of all three axes (in the Amfitrack reference frame). The following RMSE (compared to the optical reference) were observed $RMSE_x = 0.005$ m, $RMSE_y = 0.010$ m and $RMSE_z = 0.018$ m (**Right**) shows the comparison of Euclidean distance (in the Amfitrack reference frame). The following RMSE (compared to the optical reference) was observed $RMSE_{euclidean} = 0.006$ m.

Figure A.4 shows a comparison of the measured position with Amfitrack and an optical reference system during a trial where the sensor was moved in a linear motion and with a constant velocity (2, as defined in section A.2.1). In this trial

the magnetic actuator was stationary, while the sensor was manually moved approximately (without an external reference) along the x-axis. The mean euclidean distance ($RMSE_{euclidean}$ = 0.029 m, right part of Figure A.4) measured by the Amfitrack system is approximately 3 cm smaller than the optical reference. However, the measurement shows movement in all three axes ($RMSE_x$ = 0.059 m, $RMSE_y$ = 0.052 m and $RMSE_z$ = 0.167 m), while this only occurred along the x-axis. This was not the effect of a possible mismatch in the alignment of both measurement systems, as additional rotation around the z-axis results in increased errors in the relative position between actuator and sensor for the whole trial. Furthermore, for larger relative distances, more noise (with a maximum difference of 10 cm) can be observed in the magnetic measurement system.

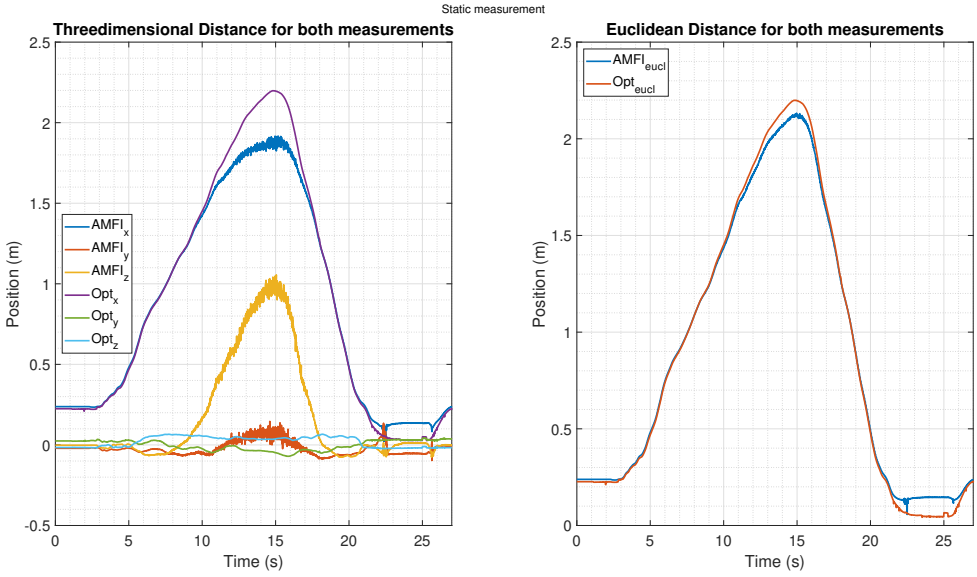


Figure A.4: During this trial the magnetic sensor was moved in a linear motion (with constant velocity) in direction of the x-axis. Position of the magnetic sensor relative to the actuator measured with both Amfitrack (AMFI) and Vicon (Opt). The following RMSE (compared to the optical reference) were observed $RMSE_x = 0.059$ m, $RMSE_y = 0.052$ m and $RMSE_z = 0.167$ m. (**Left**) shows the comparison of all three axes (in the Amfitrack reference frame). (**Right**) shows the comparison of Euclidean distance (in the Amfitrack reference frame). The following RMSE (compared to the optical reference) was observed $RMSE_{euclidean} = 0.029$ m.

Subsequent to these measurements, the Amfitrack system was strapped to the subject's body as was shown in Figure A.1(b) (described in 3 in section A.2.1) to evaluate the accuracy of a potential body-worn setup. Results of this measurement are shown in Figure A.5, with the three-dimensional relative positions on the left ($RMSE_x$ = 0.084 m, $RMSE_y$ = 0.053 m and $RMSE_z$ = 0.085 m) and the Euclidean distance on the right ($RMSE_{euclidean} = 0.020$ m). Mean errors for this trial are comparable to differences observed in the constant velocity trial with movement along one axis. However, no noise was observed here since the relative distance was smaller.

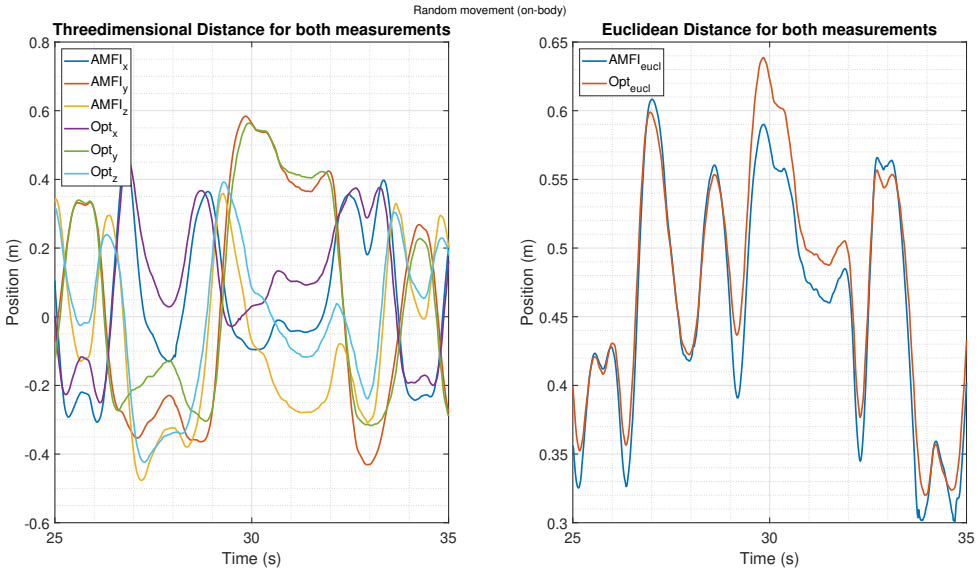


Figure A.5: During this trial the magnetic sensor was moved randomly in the on-body configuration. Position of the magnetic sensor relative to the actuator measured with both Amfitrack (AMFI) and Vicon (Opt). (**Left**) shows the comparison of all three axes (in the Amfitrack reference frame). The following RMSE (compared to the optical reference) were observed $RMSE_x = 0.084$ m, $RMSE_y = 0.053$ m, $RMSE_z = 0.085$ m. (**Right**) shows the comparison of Euclidean distance (in the Amfitrack reference frame). The following RMSE (compared to the optical reference) was observed $RMSE_{euclidean} = 0.020$ m.

A.4 Discussion

This work has shown a comparison between a magnetic measurement system and an optical reference system. This was evaluated in three conditions: static, constant velocity and on-body with variable arm movements.

The mean distance error of the Amfitrack measurement system for all conditions is approximately 3 cm (Euclidean distance), which is likely as important as errors in the different directions (x, y, z). This is an acceptable accuracy for a training application of reaching tasks according to experts in the field. Furthermore, a small relative position error is also allowed if it is consistent with the proprioceptive information of the orientation and position of the virtual hand, e.g., scaling of the virtual environment can be applied to improve this experience of the user [147]. However, this measurement system has some limitations, such as measurement range (between actuator and sensor), susceptibility to magnetic disturbances and requirement of an additional sensor to measure the orientation of the magnetic actuator (aligning the relative coordinates of Amfitrack and the virtual world). Advantages of this system are that it does not suffer from occlusions and no external computing power is required for processing of the data.

The measurement range was found to be approximately 100 cm, as can be seen by the increasing mean distance error and noise at larger distances in the constant velocity trial. Additionally, this introduces errors along different axes, while no movement along those axes occurred. This might be explained by the increased

noise of the magnetic measurement at larger relative distances. However, this maximum distance of 100 cm is sufficient for upper-body relative position measurements (magnetic actuator placed on the torso and a sensor placed on the hand). However, this is not sufficient for full-body measurements with a single actuator when both actuator and sensors are placed on-body, e.g., actuator on the pelvis and sensors on hands and feet. This could be overcome by using a larger magnetic actuator, but this is less convenient for the user because of the larger size, weight and power consumption. An alternative option might be to include an additional actuator on a different location, this can only be applied in case of a magnetic actuator as it requires time/frequency-division multiplexing. Magnetic disturbance was shown to be affecting the position accuracy of the Amfitrack system. However, the impact is minimal unless the disturbance is in between the magnetic actuator and the sensor, which is consistent with the results of Schepers et al. [169].

In the on-body configuration the sensor was placed such that a steady relative position measure was obtained. However, the trunk movement can be a relevant outcome for rehabilitation of stroke survivors. This can be measured by including an additional sensor on the chest. Furthermore, the sensor orientation is not taken into account currently, since it is not required for the proposed application because on the end-effector position is used. This information can be obtained from the measurement system and can potentially provide additional information to clinicians about reaching movements of the users. This would require evaluation of the orientation accuracy of the Amfitrack system.

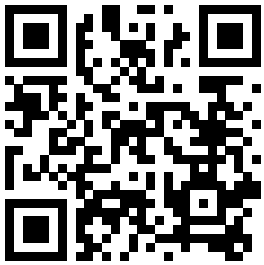
Position of the sensor is relative to the magnetic actuator, therefore, orientation of the actuator should be known to obtain positions in the global reference frame, which is mandatory for applications that use this frame. Obtaining orientation of the actuator is difficult with an inertial magnetic measurement unit (IMMU), since the orientation estimate is disturbed by the magnetic field of the actuator. An IMMU uses the magnetometer information to obtain an estimate of the heading, which is important to align the global coordinate frame of the virtual environment with the measured relative positions, such that it functions for all body orientations, i.e., when the user looks in different directions. This also was a limitation for the developed VR application mock-up, which functions only if the magnetic actuator is in the correct orientation because of the lacking integrated IMMU (in the current setup), which should be improved in future work.

A.5 Conclusion

Accuracy of the Amfitrack measurement system ($\sim$3 cm) is promising for use in a VR application for training reaching tasks of stroke survivors. Additional research is required for evaluation of the suitability of the proposed VR environment for training exercises.

A.6 Supplementary Material

A visualization of the concept VR environment (with HTC Vive control) can be found at `https://youtu.be/ph6BMBS-AXs`:



A visualization of the concept VR environment (with Amfitrack control) can be found at `https://youtu.be/_YGJmZ9kQmg`:

Acknowledgements

The moment that all the research and writing is finished has finally come. For just over four years I have been in the roller coaster, also called a PhD, during which I have learned so much (both academically and personally). However, this would not have been possible without support and guidance from the people around me. Therefore, I would like to show my thanks to those people in this section, sincerely hoping not to forgot to mention anyone (my apologies if I do).

First, I would like to express my deepest appreciation to my supervisor prof. dr. ir. Peter H. Veltink, who provided me with extensive support and interesting discussions for the duration (and slightly beyond) of my doctoral research. Without your help and expert knowledge in the field of ambulatory human motion sensing, this work would not have been possible.

Dr. ir. Matteo Giuberti was instrumental in defining the course of this research. For this, I am extremely grateful. Your extensive feedback on paper drafts really helped me to improve my academic writing skills. I already miss our lengthy discussions, while working at Xsens.

I would like to express my sincere gratitude to dr. ir. Bert-Jan F. van Beijnum, for the extensive guidance in my doctoral research. Your knowledge of machine learning in combination with human motion has been a great source of inspiration. And your positive attitude helped me to overcome hurdles in the trajectory.

Special thanks to prof. dr. ir. Mannes Poel for all the machine learning guidance, which really helped me get up to speed quickly in the first part of my PhD. And to whom I could always turn whenever I got stuck.

I would like to thank my external supervisor dr. ir. Henk Luinge, for guidance during my first year as a PhD candidate. And for allowing me to work twice a week from an office in Xsens, which helped me to get feedback on issues quickly.

Dr. ir. Giovanni Bellusci, I am grateful that you took over Henk's role as external supervisor and provided continuation to the great support from Xsens. And eventually for facilitating continuation to the research, by hiring me as an Xsens employee.

Chapters 4 and 6 are the product of a fruitful collaboration with Roessingh Research & Development. Many thanks to Eric Maartens and Jasper Reenalda for the biomechanical insights and invaluable help with the measurements. In particular I would like to express my sincere gratitude to Leendert Schaake, who not only provided me with excellent measurement equipment and space, but also provided extensive support in setting up and processing basically all the measurements in this work.

Appendix A is the product of a worthwhile collaboration with Roessingh Research & Development and RE-liON. I gratefully acknowledge the assistance of prof.

dr. ir. Jaap Buurke for the design of a virtual reality application to train reaching movements of stroke survivors. I would like to extend my sincere thanks to Marin Jonkman for the support with the provided magnetic measurement system. Additionally, I would like to thank Daan Nusman for his contribution to the project.

Chapter 5 was possible with the help of bachelor student Stephan Jaspar, who designed and tested the initial algorithms, so thank you for your significant contribution. Furthermore, I would like to extend my thanks to Jaap Harlaar of VUmc for his biomechanical contribution to the project and this study in particular.

During my PhD I had the opportunity to work both at the Biomedical Signals & Systems group (BSS, University of Twente) and at Xsens. This allowed me to work in both an academic and professional environment at the same time, and meet and learn from many more experts in their respective fields.

From the Xsens side, I would like to express my gratitude to Nina for her support during the final few months. And to Michele, Fabian, Jason and many others with whom I had interesting discussions and ping-pong matches. I am grateful that I can now call you all my colleagues.

From the BSS side, I would first like to express my gratitude to Wies and Sandra, both secretaries, who allowed me to focus on the research side of the PhD. Furthermore, Marcel and Ed, thanks to your great technical support I could never blame the equipment for anything. And of course, all my office mates, Kees, Frauke, Fokke, Angelos, Tom, Kostas, Stefan and Roelof for making this PhD trajectory a fun and worthwhile time. And of course all the other colleagues at BSS, thank you for providing a truly multi-disciplinary research environment. Last but not least, thanks for the amazing work and inspiration for new directions in my research, Sander, Stephan and Klaas-Jan.

Angelos and Kostas, I am very grateful that you two awesome Greeks accepted to support me during my defense as my paranymphs. Angelos, thanks for sharing an office with me for the largest part of my PhD and for inspiring and supporting me with our many interesting (Dutch) discussions. Kostas, thanks for providing the best office music, interesting machine learning discussions and great trips to enjoy house music.

I would like to thank my family for allowing me to start this adventure and providing the necessary support and encouragement during the whole PhD period.

And last, but most certainly not least, I am eternally grateful to my wife Marieke for her love and endless support, you kept me sane during this trajectory. Thank you for helping me finish my PhD thesis. And Liam, thank you for being the inspiration and distraction needed to finish this work in a timely manner.

Biography

Frank Wouda was born in Hengelo (O), the Netherlands, on the 7th of November 1990. He finished his VWO at the Waerdenborch, Holten, 2009.

After his VWO he started studying Mechanical Engineering at the University of Twente, Enschede, the Netherlands. In 2014 he did a four-month internship at Össur (Reykjavik, Iceland) where he used motion capture technologies for the first time. The topic of his Master's thesis was: the design and evaluation of a balance controller during human gait, which was based on the energy of an inverted pendulum.

In May 2015 he started his PhD project, MiniSens (funded by STW), at the Biomedical Signals and Systems group of the University of Twente. The focus of this project was to reduce the number of sensors for full-body motion capture. He worked on the development of data-driven approaches for minimal sensing and applying these to specific applications.

Since August 2019, Frank is employed by Xsens Technologies as a research engineer on topics related to inertial motion capture, machine learning and applications.

List of Publications

Published peer-reviewed journal publications

F. J. Wouda, M. Giuberti, G. Bellusci and P. H. Veltink, "Estimation of Full-Body Poses Using Only Five Inertial Sensors: An Eager or Lazy Learning Approach?" *Sensors*, vol. 16, no. 12, 2016.
https://doi.org/10.3390/s16122138

F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. J. F. van Beijnum and P. H. Veltink, "Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors" *Frontiers in Physiology*, vol. 9, no. 218, 2018.
https://doi.org/10.3389/fphys.2018.00218

F. J. Wouda, M. Giuberti, N. Rudigkeit, B. J. F. van Beijnum, M. Poel and P. H. Veltink, "Time Coherent Full-Body Poses Estimated Using Only Five Inertial Sensors: Deep versus Shallow Learning" *Sensors*, vol. 19, no. 17, 2019.
https://doi.org/10.3390/s19173716

Submitted peer-reviewed journal publications

F. J. Wouda, S. L. J. O. Jaspar, J. Harlaar, B. J. F. van Beijnum and P. H. Veltink, "Foot Progression Angle Estimation Using A Single Foot-Worn Inertial Sensor"

Conference contributions

F. J. Wouda, M. Giuberti, G. Bellusci and P. H. Veltink, "Estimation of Full-Body Poses Using Only Five Inertial Sensors: An Eager or Lazy Learning Approach?" *6th Dutch Conference on Bio-Medical Engineering*, Egmond aan Zee, pp. 143, 2017.

F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. J. F. van Beijnum and P. H. Veltink, "On the Validity of Different Motion Capture Technologies for the Analysis of Running," *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Enschede, pp. 1175-1180, 2018.
https://doi.org/10.1109/BIOROB.2018.8487210

F. J. Wouda, M. Giuberti, G. Bellusci, G., van Beijnum, B. J. F. van Beijnum and P. H. Veltink, "Improving Full-Body Pose Estimation from a Small Sensor Set Using Artificial Neural Networks and a Kalman Filter" *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33(01), pp. 10063-10064, 2019.
https://doi.org/10.1609/aaai.v33i01.330110063

Pending Patent Applications

F. J. Wouda, S. L. J. O. Jaspar, B. J. F. van Beijnum and P. H. Veltink, "Foot Progression Angle Determination Method and System, Computer Program and Storage Medium"

For an updated list of publications and citations see Google Scholar: