TUM School of Computation, Information and Technology
Technische Universität München

Dissertation

Physical Activity Recognition Using Mobile Devices

Lara Marie Reimer

TUM School of Computation, Information and Technology
Technische Universität München

Physical Activity Recognition Using Mobile Devices

Lara Marie Reimer, M.Sc.

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:
Prof. Dr.-Ing. Jörg Ott

Prüfer*innen der Dissertation:
1. Prof. Dr. rer. medic. Stephan Jonas
2. Prof. Dr.-Ing. Veit St. Senner

Die Dissertation wurde am 24.06.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 05.12.2022 angenommen.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Stephan Jonas and Prof. Veit Senner. By a lucky coincidence, Prof. Jonas joined TUM just by the time when I decided to continue my academic career. Ever since, he supported me, challenged me, and enabled me to work on a lot of different and exciting projects in sports and health projects. I am truly grateful for the possibilities you offered, for your trust, support, and motivation throughout this time. Thanks to Prof. Senner for his agreement to take over the role of the second examiner and for providing feedback. I really enjoyed the contact to other researchers working on quite similar topics as mine, but from a slightly different perspective, and hope to continue with the collaborations we just started. I would also like to thank Prof. Jörg Ott, who did not only agree to take over the role of the chair of my dissertation, but also supported me and my fellow colleagues in the past year in various aspects.

Many thanks go out to my mentor, Dr. Jan Ole Johanssen. You have been my mentor for so long now, starting during my Master's thesis in 2019, that it just feels natural to share and discuss my research with you. Thank you for always having an open ear, providing feedback, and brainstorming with me.

While working on interesting projects is important, it is only a part of the whole picture of worklife. At this point, I would like to thank my Digital-Health-colleagues Marko Jovanović, Maximilian Kapsecker, Jens Klinker, Leon Nissen, and Jonas Schulte-Coerne from the bottom of my heart. With you, work often does not even feel like work - it is more like spending time with your friends.

Apart from the Digital Health Group, I would also like to express my gratitude towards all of the great people from the LS1 chair. Prof. Bruegge, thank you for inspiring and motivating me through your courses to find my passion for research. Dr. Lukas Alperowitz, Dr. Jan Philip Bernius, Dr. Dora Dzvonyar, Dr. Nadine von Frankenberg, Dr. Dominic Henze, Dr. Jan Ole Johanssen, and Dr. Paul Schmiedmayer: thank you for your friendship, for incredible days at the WWDC, and funny evenings in Munich.

Large parts of this research would not have been possible without the help of several students. First of all, I would like to thank my student research assistants Céline Aldenhoven and Florian Schweizer for being open to my ideas, contributing your own thoughts, and realizing our projects together. Moreover, I would like to thank the students who contributed through their final theses to my research: Severin Weigel, Bhagyashree Padalkar, Malintha Adikari, Florian Ehrenstorfer, Hans Santoso, Takashi Fukushima, Konstantin Kuchenmeister, Christian Brechenmacher, Céline Aldenhoven, Klaudia Madhi, Nikol Nikolova, Hannes Simon, Joshua Pusch, Annabell Vogelsang, Alicia Mazurkiewicz, Leonard Pries, Florian Schweizer, and Cameron Shemilt.

My research was also supported by other researchers, whom I would like to thank at this point. Dr. Florian Kreuzpointner, thank you for introducing me and my students to Vicon, for taking the time to explain how the system works, and to ensure that our setup delivers reliable data. Thanks also to the DigiMed AP 1.2/APP Team, especially Prof. Heribert Schunkert and Dr. Fabian Starnecker for complementing our technical knowledge with your medical expertise. Thank you, Dr. Wolfgang Birkle, for introducing me to the pretty complex and interesting world of golf.

As interesting as the world of research might be, it still contains a large amount of bureaucracy and administrative tasks. At this point, I would like to thank everyone who supported me with these tasks: Ruth Demmel, Christine Lissner, Monika Markl, Helma Schneider, and Petra Weiss, but also the LS1 admin team.

I would like to express my deepest gratitude towards my family, my parents Thomas and Monika, my brother Lukas, but also those, who cannot be with us any more. You were always there for me, enabled me to do whatever I wanted, and supported me in every aspect of my life, always wanting the best for me. I could not imagine doing this without you. Thank you!

Last, but not least - thanks to my boyfriend, Dr. Jan Philip Bernius. Thank you for being there at every step of my/our way towards our dissertation(s), thank you for encouraging me, for motivating me, for just being there in good and in bad moments ever since we met. It really means a lot to me to have you by my side.

This work was supported by a grant from Software Campus through the German Federal Ministry of Research and Education, grant number 01IS17049. Parts of this research were funded by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de).

Abstract

The usage of mobile devices in the daily life of people is increasing, and so is the potential for utilizing these devices to improve people's lives. People can monitor their overall health data by using more integrated sensors, like their heart rate or physical activity. Current solutions often focus on recognizing specific activities but fail to provide feedback on how well the activity was performed. Detailed motion analysis is still almost exclusively carried out in laboratories and comes with high costs, limiting its application and availability. The demand for affordable and accessible solutions to self-analyze physical activity in health and fitness applications is increasing. Recent developments in smartphones, integrated processing units, and sensors enable new possibilities for physical activity recognition and analysis using mobile devices.

This dissertation explores the possibilities of physical activity recognition and analysis using mobile devices in mobile health and fitness applications. It provides an overview of software frameworks for mobile motion capture, evaluates the accuracy of the 3D motion capture framework Apple *ARKit*, and delivers insights into how these frameworks can be applied to several use cases in health and fitness.

This dissertation provides evidence that the *ARKit* framework accurately tracks the course of motion. However, the results exhibited significant deviations of the measured joint angles compared to the *Vicon* system, a gold standard for motion capture. The accuracy of the tracking results mainly depended on the visibility of the tracked joints and the motion performed. Five case studies showed that physical activity recognition using mobile devices could be applied to perform activity classification (Case Study 1), exercise counting (Case Study 2), and balance assessments (Case Study 4). The prototypical applications achieved good usability results (Case Study 4) and good user acceptance among one of the target groups (Case Study 3). Case Study 5 provided an approach to leveraging health data, including physical activity data, for scientific studies while preserving high data privacy standards.

Zusammenfassung

Die Nutzung mobiler Geräte im Alltag der Menschen steigt, und damit steigt auch das Potential, diese Geräte im Alltag einzusetzen, um das Leben der Menschen zu verbessern. Durch die Nutzung integrierter Sensoren können die Menschen allgemeine Gesundheitsparameter wie ihre Herzfrequenz oder körperliche Aktivität überwachen. Aktuelle Lösungen fokussieren dabei häufig die Erkennung bestimmter Aktivitäten, aber bewerten nicht, wie gut die Aktivität ausgeführt wurde. Eine detaillierte Bewegungsanalyse wird weiterhin fast ausschließlich in Laboren durchgeführt und ist mit hohen Kosten verbunden, was ihre Anwendung und Verfügbarkeit limitiert. Die Nachfrage nach bezahlbaren und verfügbaren Lösungen für eine selbstständige Bewegungsanalyse in Gesundheits- und Fitnessanwendungen steigt. Die neuesten Entwicklungen in Smartphones, integrierten Prozessoren und Sensoren ermöglichen neue Möglichkeiten der Bewegungserkennung und -analyse mit Mobilgeräten.

Diese Dissertation erforscht die Möglichkeiten der Erkennung und Analyse körperlicher Aktivität mit Mobilgeräten in Gesundheits- und Fitnessapplikationen. Sie präsentiert einen Überblick über Software Frameworks für die mobile Bewegungserkennung, evaluiert die Genauigkeit des 3D-Frameworks Apple *ARKit* und zeigt Einblicke, wie diese Frameworks in Anwendungsfällen im Gesundheits- und Fitnesskontext eingesetzt werden können.

Die Dissertation zeigt, dass das *ARKit* Framework den Bewegungsverlauf mit hoher Genauigkeit erkennt. Allerdings weisen die Ergebnisse signifikante Abweichungen der Gelenkwinkel zum *Vicon* System auf, einem Gold-Standard-System für Bewegungserkennung. Die Genauigkeit der Bewegungserkennung hängt insbesondere von der Sichtbarkeit der verfolgten Gelenke und der durchgeführten Bewegung ab. In fünf Fallstudien konnte gezeigt werden, dass die Erkennung von Körperbewegungen mit Mobilgeräten zur Aktivitätsklassifikation (Fallstudie 1), zur Zählung von Übungswiederholungen (Fallstudie 2) und zur Durchführung von Balancetests (Fallstudie 4) verwendet werden kann. Die prototypischen Applikationen erreichten gute Nutzbarkeitsergebnisse (Fallstudie 4) und Akzeptanz bei einer der Zielgruppen (Fallstudie 3). Fallstudie 5 präsentiert einen Ansatz, über den Gesundheitsdaten, wie beispielsweise Aktivitätsdaten, für wissenschaftliche Studien unter Wahrung hoher Datenschutzstandards genutzt werden können.

Publication Preface

The contribution of this dissertation is based on the following three peer-reviewed lead-/main author publications in international conferences and journals:

Publication [A]

Lara Marie Reimer, Severin Weigel, Florian Ehrenstorfer, Malintha Adikari, Wolfgang Birkle, Stephan Jonas
Mobile Motion Tracking for Disease Prevention and Rehabilitation Using Apple ARKit
dHealth 2021 - Navigating Healthcare Through Challenging Times

DOI: 10.3233/SHTI210092

Publication [B]

Lara Marie Reimer, Maximilian Kapsecker, Takashi Fukushima, and Stephan M. Jonas
Evaluating 3D Human Motion Capture on Mobile Devices
Applied Sciences - Special Issue on Applied Biomechanics and Motion Analysis

DOI: 10.3390/app12104806

A data set was published alongside publication B, containing the data gathered during the laboratory experiment described in the publication:

Lara Marie Reimer, Maximilian Kapsecker, Takashi Fukushima, and Stephan M. Jonas
Evaluating 3D Human Motion Capture using Apple ARKit against the Vicon System:

A Dataset
Zenodo

DOI: 10.5281/zenodo.6527283

Publication [C]

Céline Madeleine Aldenhoven, Lara Marie Reimer, and Stephan Jonas
mBalance: Detect Postural Imbalance with Mobile Devices
dHealth 2022

DOI: 10.3233/SHTI220344

The following publication is embedded for additional context and has not been considered as part of the assessment of this dissertation:

Publication [D]

Lara Marie Reimer, Fabian Starnecker, Heribert Schunkert, and Stephan Jonas
Developing an App for Cardiovascular Prevention and Scientific Data Collection
dHealth 2021 - Navigating Healthcare Through Challenging Times

DOI: 10.3233/SHTI210095

Contents

1 Introduction

Regular physical activity is associated with positive influence on functional capacity and mood [1, 2]. It reduces the risk of Cardiovascular Disease (CVD) [1], diabetes mellitus type 2, and dementia [3]. Physical activity is also beneficial during oncologic and hematologic treatments to counteract muscular atrophy and improve the patients' mood [4–6]. Apart from these benefits, physical activity and human motion in general exhibit relevance across various use cases in health and sports. The presence of specific motion patterns or abnormalities can be used to recognize particular exercise types, measure adherence or exercise regimen, hint at underlying health conditions, or indicate problematic movements that could lead to injuries.

Human Motion Capture (HMC) and analysis deals with detailed tracking and analysis of human motion. It has been a field of interest since the 1970s [7] and has already enabled various industry applications [8], sport and kinematics applications [9, 10], and health and medical applications [11–13]. Traditionally, HMC has been performed using specialized motion capture systems in laboratories, which are still commonly used for HMC. However, recent advances in consumer-grade devices such as smartphones, tablets, and smartwatches enable HMC on mobile devices [14–17]. These devices open up new possibilities in sports and health contexts, such as detecting specific activities and monitoring biomarkers. However, HMC is still a field of active research. This dissertation targets some of the open topics.

1.1 Problems in the Current Application of HMC

With the ubiquity of smart devices such as smartphones and -watches, the opportunities for health and fitness applications has increased, too. Using a consumer-grade device in a health and fitness context is linked to several problems:

(1) Traditional HMC approaches are expensive.

In the health and medical area, the physicians or other trained staff usually perform motion assessments as part of examinations. Technology-based methods in a laboratory are rarely used due to their high costs [18] and are limited to specific use cases such as gait analysis [19].

Motion quality analysis in sports and fitness is mainly reserved for two use cases: professional athletes who want to improve their performance, or amateur athletes, who experience pain during physical activity and want to investigate its cause. Traditional personal coaching through qualified trainers is widely used and can be complemented by technology-assisted approaches.

In both areas, HMC requires a trained rater and causes high additional costs for the acquisition of HMC technology and the operation of such systems.

(2) HMC is often performed irregularly or in extensive intervals.

Motion analysis is often only performed in lengthy intervals. In health use cases, these intervals often span several months or years between assessments. The assessments are typically part of examinations by a physician. Especially with neurodegenerative diseases, motion quality can vary between days [20]. These variations leave the physician with a potentially incorrect image of the current state of their patient, as they might see their patient on an extraordinarily good or bad day. Moreover, the assessment depends on the observing rater and can vary between raters, which depicts an additional problem in HMC.

In sports, assessments are done more regularly in intervals such as weeks or months. The assessments usually take place during training sessions with a professional sports coach. In practice, the assessment is often performed manually. In sports, regular HMC and feedback on the performed motion are crucial to identify errors in the technique, prevent injuries and increase the performance [21, 22].

(3) HMC using mobile devices is still immature.

Mobile devices such as smartphones and smart wearables enable a variety of mobile health and fitness applications that are already available on the market. Health and fitness applications are used by a large amount of the people, especially in younger generations [23, 24], with increasing tendency [25, 26]. They help to reach their users' health and fitness goals [23, 27] and have a positive impact on their life [28].

Mobile devices can use different technologies to capture motion. The most common systems used on consumer-grade smart devices are Inertial Measurement Systems (IMSs) and Image Processing Systems (IPSs).

However, existing HMC technologies and applications are connected to four sub-problems:

(3a) While several applications exist for specific use cases, many use cases are not covered by mobile applications.

Most applications in the health sector mainly include features informing about specific health parameters, such as the diet or the amount of physical activity, and often act as life coaches [23]. Applications targeting motion analysis focus on specific use cases, such as the quantification of fall risks or the detection of falls [29–38], while others are not covered at all.

(3b) IMS-based applications still rely on several sensors to assess motion quality, limiting their availability and accessibility.

Mobile IMSs rely on Inertial Measurement Units (IMUs) to detect the acceleration and rotation of the device containing the IMU. In physical activity recognition, IMUs can be attached to one or more body parts.

Most smart devices contain IMUs, which are used to detect specific physical activity. Depending on where the IMU is placed, physical activity data can be obtained, such as the hip rotation when the smartphone is positioned in the pants' pocket. This functionality is commonly used by smart devices, such as for step counting. Such activity data is often tracked by mobile health and fitness applications. More advanced solutions allow tracking of workouts, such as running and capture additional parameters like cadence, speed, heart rate, or calories burnt. However, these applications rarely provide feedback, especially regarding the motion quality [39].

Research has shown that IMSs can detect and classify movements and are thus a suitable option for mobile motion capture of human physical activity [40, 41]. Mobile IMSs can measure joint angles if attached to the joints and can thus provide feedback on the motion quality of this joint [32, 33]. However, if multiple joints need to be tracked, several IMU sensors need to be used, which decreases the accessibility and availability of such systems.

(3c) IPS-based applications have only been evaluated to a limited extent.

Mobile IPSs use optical sensors such as cameras to track objects and can be complemented by additional depth sensors like Light Detection And Ranging (LiDAR) and structured light sensors. Mobile optical motion capture requires a camera, sufficient lighting conditions, and sufficient processing power to run the computer vision analysis. Both camera and processing units are integrated into newer smartphones, so no additional equipment is needed to perform motion analysis. As they can track the whole body without additional sensors, they represent a promising solution for affordable and accessible HMC.

Several studies have shown that mobile IPS solutions can track human motion with good accuracy [42–45]. The accuracy of the results mainly depends on the performed task, the video quality, and the position of the recording device. Especially the camera position is a relevant factor, which is a common problem of motion capture performed from only one camera perspective [46]. Recently, 2D frameworks have been complemented by 3D frameworks, which promise to overcome the limitations of the camera perspective by detecting the body in 3D [47–51]. However, little research has been performed on the accuracy of mobile 3D motion capture frameworks, especially in health and fitness applications.

(3d) Mobile applications often provide insufficient data security and privacy.

Reviews by Krebs et al. and Martinez-Perez showed that insufficient user data privacy is among the main reasons users do not use such applications [24, 52]. Especially mobile applications that store health information need to protect data from manipulation and unauthorized access and have a high potential of damage [53].

1.2 Solution

The described problems (1) and (2) highlight a demand for accessible and affordable solutions to perform self-analysis of physical activity to deliver regular feedback, decrease the risk of injuries, and increase the performance of (amateur) athletes. Such solutions would allow more detailed and objective monitoring and help identify possible problems early. As such solutions handle sensitive user data, they must adhere to high data security and privacy standards.

A promising approach to these problems are mobile motion capture systems. Mobile motion capture systems describe motion capture technologies characterized by their high degree of mobility and especially include devices such as smartphones, tablets, and other consumer-grade wearables, which are already used by over a billion of people worldwide [54]. They represent accessible technologies and are available at low costs compared to traditional motion capture systems, thus able to solve problems (1) and (2). Moreover, most mobile devices provide open-development platforms and thus allow the development and installation of third-party applications, enabling developers to build and distribute their own applications. These possibilities open up the development of applications targeting new use cases in health and fitness contexts, solving problem (3a).

In this dissertation, we therefore leverage the possibilities of mobile motion capture in the context of physical activity recognition for mobile health and fitness applications. The problems (3b-d) depict open questions in the field of mobile HMC. Due to the limitations of mobile IMSs (problem (3b)), we apply mobile IPSs for physical activity recognition and analysis, especially if a motion quality assessment needs to be performed. To target problem (3c), we validated the mobile IPS technology Apple *ARKit* for health and fitness applications. Additionally, we described and evaluated different use cases for physical activity recognition and analysis (problems (3a)) as well as a possibility for adhering to high data privacy standards for scientific data collection (problem (3d)).

1.3 Objective

In the light of the problems mentioned above, the objective of this dissertation is threefold and defined by the RQs:

- **RQ 1:** Which possibilities for mobile optical motion capture exist?
- **RQ 2:** How do optical mobile motion capture systems compare to established solutions?
- **RQ 3:** In which fitness, medical, or health-related use cases can optical mobile motion capture systems be applied?

To answer RQ 1, the current state-of-the-art of research regarding the possibilities and limitations of HMC on mobile devices in the context of physical activity recognition is evaluated. As IMU-based systems exhibit limitations regarding accessibility and availability for mobile motion capture, this dissertation focuses on investigating mobile IPSs for motion quality analysis in health and fitness applications.

Second, mobile IPS systems for applications in the sports and health sector are validated, targeting RQ 2. The validation delivers insights into the motion capture accuracy of mobile IPSs including their strengths and weaknesses. The validation of mobile IPSs is essential to apply them in the context of health and fitness.

Third, possible use cases of physical activity recognition using mobile devices in sports and health applications are explored, focusing on motion quality analysis (RQ 3). The use cases are analyzed regarding their technical feasibility, perceived usefulness among target groups, and usability. These results draw further conclusions on the potential of physical activity recognition using mobile devices.

2 Research Process and Methods

Within this dissertation, we explored the possibilities of physical activity recognition and analysis using mobile devices, specifically in sports and health applications, as already defined in Section 1.3. The dissertation is publication-based and thus includes contributions based on several publications, which are summarized and set into context by this document. This chapter describes the dissertation's research process, the methods used in the publications, and the relations between the publications. Figure 2.1 visualizes the research approach, including the three Research Questions (RQs), their subquestions, and the applied methods.

2.1 RQ 1: Which possibilities for mobile optical motion capture exist?

Recent literature on mobile motion capture technologies, especially using mobile devices, was reviewed. The literature review aimed to get an overview of the field's state-of-the-art, investigate the strengths and weaknesses of the currently available systems, and identify related work. The research focused on three main questions:

- **RQ 1.1:** How is Human Motion Capture (HMC) defined, and how is it performed?
- **RQ 1.2:** How can HMC be performed on mobile devices?
- **RQ 1.3:** Which possibilities exist for mobile optical motion capture using markerless methods?

The literature research was based on the digital libraries of IEEE Xplore, ACM Digital Library, and PubMed. We compared different HMC technologies and their suitability for the objective of this research, particularly of mobile Image Processing System (IPS) technologies.

As the main result of the literature review, we selected Apple *ARKit* as the core technology investigated within this research. *ARKit* is a mobile 3D motion capture framework using optical sensors. It was selected due to its good embedding into the iOS

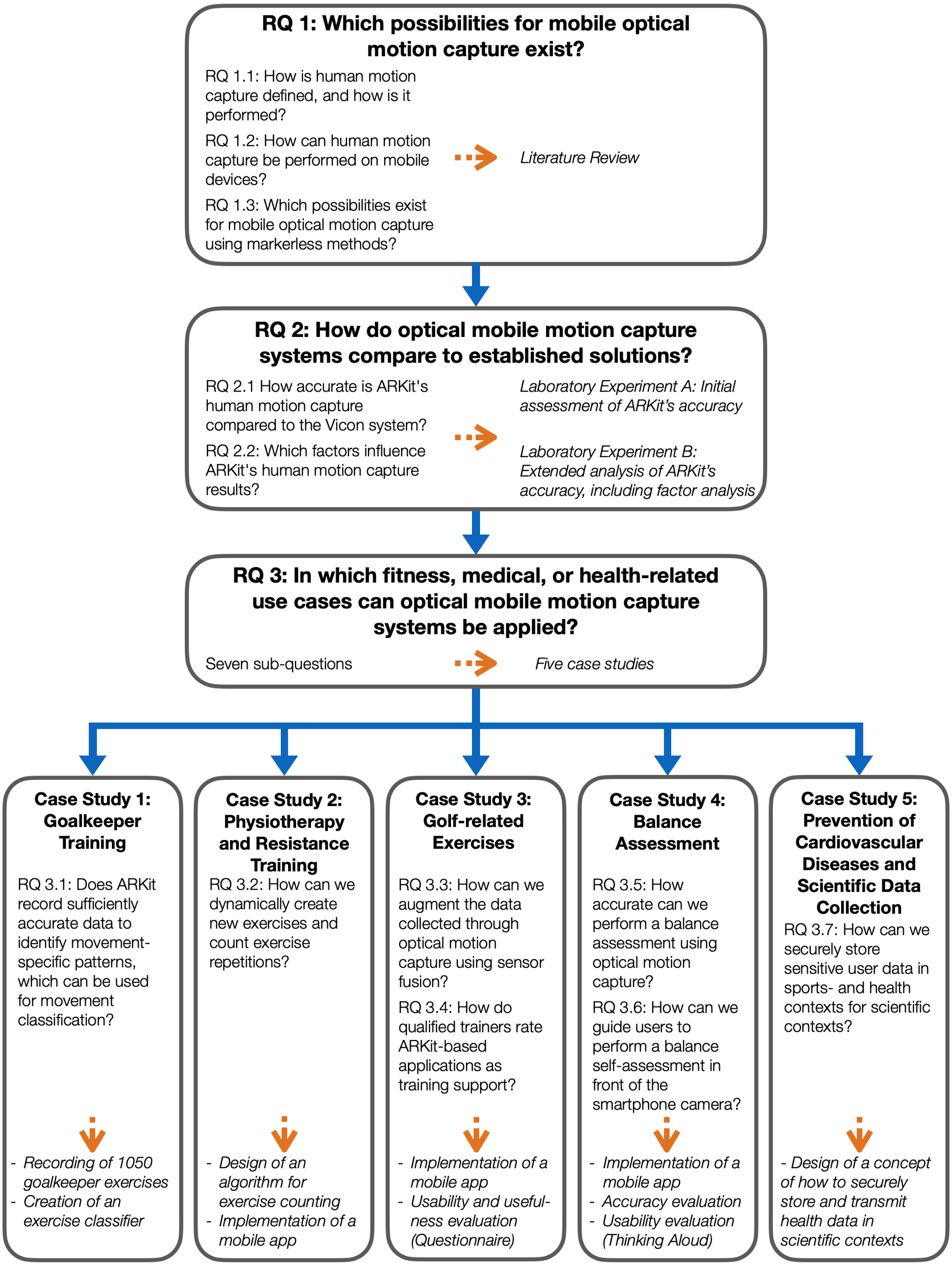


Figure 2.1: Overview of the research approach.

operating system, enabling it to use additional smartphone sensors such as the integrated IMU and depth sensors. With the advances in Apple's M1 processor, *ARKit* uses a variety of sensors combined with high processing power, which is a unique combination in current mobile IPS software. Despite these advantages, *ARKit* has only been evaluated to a limited extent. The results of the literature research are summarized in Chapter 3.

2.2 RQ 2: How do optical mobile motion capture systems compare to established solutions?

RQ2 was separated into two sub-questions to achieve better differentiation:

- **RQ 2.1:** How accurate is *ARKit*'s Human Motion Capture compared to the *Vicon* system?
- **RQ 2.2:** Which factors influence *ARKit*'s Human Motion Capture results?

Two laboratory experiments (A & B) followed by quantitative data analysis were performed to answer RQ 2.1 and RQ 2.2. In both experiments, we validated the Apple *ARKit* framework against the *Vicon* system, which is a gold standard for motion capture. In each experiment, the subjects were asked to perform physical exercises while being recorded by the two systems. Subsequently, the data was compared to assess the accuracy of *ARKit*'s motion capture. In the extended analysis performed in B, potential influencing factors on the accuracy of *ARKit* were studied. This section summarizes the used methods of Publications [A] (RQ 1) and [B]. More detailed descriptions including the results are described in Chapters 4 and 5.

2.2.1 Laboratory Experiment A

The goal of laboratory experiment A (Publication [A], RQ 1 [55]) was to gain an initial understanding of the accuracy of the data recorded by *ARKit*. This included the ability to track specific movements at all, the overall accuracy, and possible limitations. Laboratory experiment A included 12 subjects, five females and seven males. The subjects were tracked by an iPad Pro (11"2020, LiDAR, Apple Inc., Cupertino, CA, USA) at a distance of $3m$ from a frontal position. Simultaneously, a 10-camera *Vicon* setup using the full-body Plug-in-Gait model [56] was used to record the ground truth data of the body motion (Figure 2.2). *ARKit* sampled with a variable sampling rate of approximately $60Hz$, *Vicon* with $250Hz$. During the experiments, the subjects were asked to perform a set of nine predefined activities focusing on the upper and lower body at different speeds and fast, abrupt changes of speed. The activities included walking and running on a treadmill at 3, 5, and $7 km/h$, passing and catching a ball, jumping jacks, cuttings, and squats.

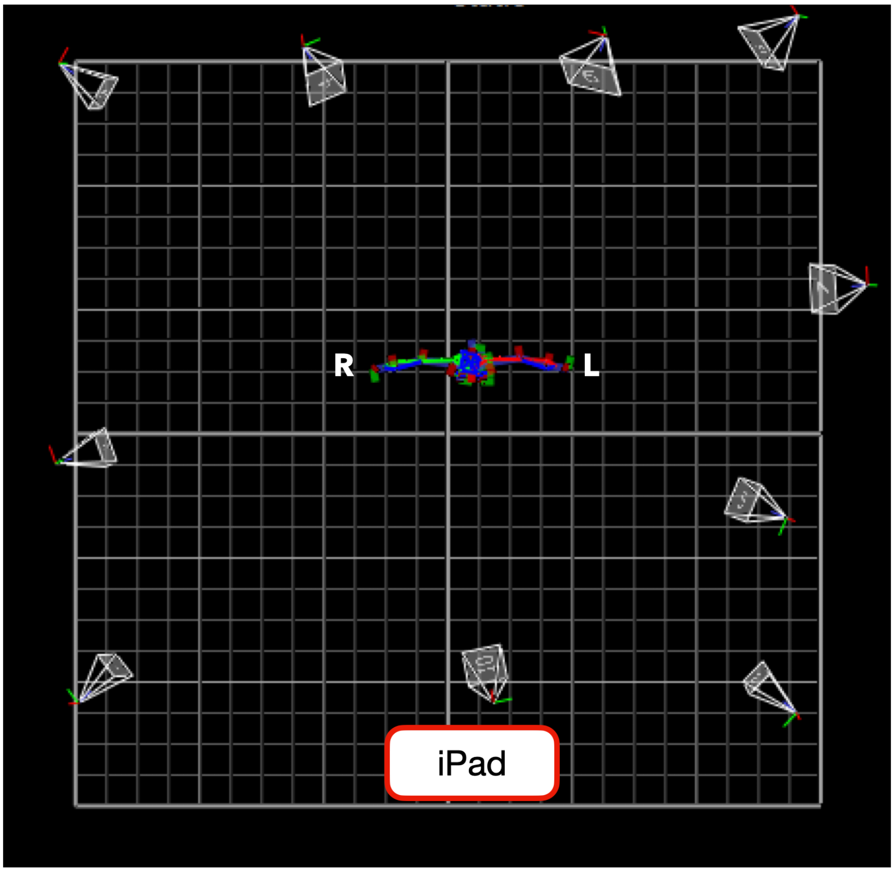


Figure 2.2: Camera setup of the laboratory experiment A. The subject is positioned at the center of the tracking area. 10 infrared cameras of the *Vicon* system are positioned around the subject. An iPad is positioned in front of the subject. "L" and "R" mark the subject's left and right body side, the subject is facing the iPad.

We calculated both systems' minimum, maximum, mean, standard deviation, and range of motion for neck, shoulder, elbow, knee, and ankle angles. These metrics were chosen to assess whether *ARKit* is able to track the full range of motion during the exercises or whether it is limited to a specific range and cannot detect joint angles below or above a certain threshold. Additionally, the Pearson correlation coefficient was calculated to quantify how well *ARKit* can track the course of the motion.

2.2.2 Laboratory Experiment B

Laboratory experiment B (Publication [B] [57]) was performed to deepen insights into *ARKit*'s accuracy and by investigating which factors possibly influence the tracking accuracy. The study design was based on the results of laboratory experiment A. Laboratory experiment B included ten subjects. To investigate *ARKit*'s performance in different viewing angles, two iPad Pros (11" 2021, LiDAR, Apple Inc. Cupertino, CA, USA) were placed, one in a frontal position, the other at a 30° angle to the side. The setup of the *Vicon* system consisted of 14 infrared cameras. Again, the full-body Plug-in-gait model of the *Vicon* Nexus software captured the ground truth data [56]. Similar to laboratory experiment A, *ARKit* recorded at a variable sampling frequency of around $60Hz$, *Vicon* at $250Hz$. The subjects performed a set of eight body-weight exercises, targeting the whole body. Based on the previous findings, the exercises were selected such that both *ARKit* and *Vicon* could fully track the motion. This mainly excluded exercises where more extensive proximal parts of the body were positioned near the floor, such as push-ups. The resulting exercises included Squats, Front Lunges, Single Leg Deadlifts, Side Squats, Lateral Arm Raises, Reverse Flies, Jumping Jacks, and Leg Extension Crunches. Each exercise was performed for approximately ten repetitions. Some subjects performed an extra repetition, which was also included in the resulting data set.

During the data analysis, the data sets from both iPads were compared against the *Vicon* data set, which was downsampled to match *ARKit*'s recording frequency. Using cross-correlation, the recordings were aligned. Subsequently, metrics were calculated to assess *ARKit*'s accuracy for all viewing angles, exercises, joint angles, and subjects, resulting in 1048 exercise/angle/subject/view combinations. The weighted Mean Absolute Error (wMAE) and the mean Spearman Rank Correlation Coefficient (SRCC) were used as primary indicators for the accuracy. The wMAE was selected to identify the mean absolute deviation between *ARKit* and *Vicon* for each exercise/angle/subject/view combination and used to identify, which combinations resulted in higher or lower errors. Similarly, the mean SRCC was calculated for each of the combinations and used to conclude in which combination *ARKit* was able to accurately follow the joint motion. The wMAE and SRCC were compared to generate additional insights into the data, such as whether certain combinations existed, where *ARKit* exposed higher wMAE values, but good SRCC values. Such a combination would indicate that *ARKit* successfully tracks the motion, but fails at detecting the full range of motion of a joint in a specific exercise. To further evaluate the impact of possible influencing factors on the accuracy, factor analysis was performed using Welch ANOVA, t-test, and logistic regression on

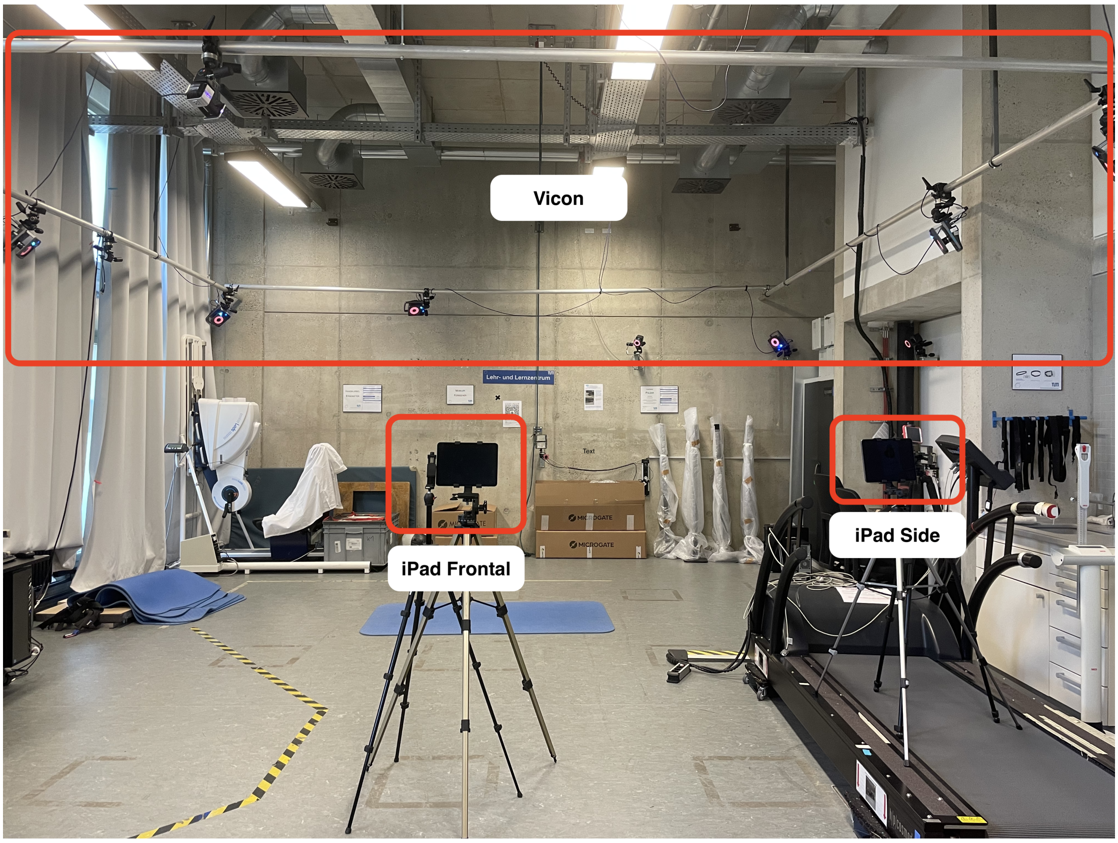


Figure 2.3: Camera setup of laboratory experiment B. The 14 *Vicon* infrared cameras are positioned around the subject close to the ceiling of the laboratory. Two iPads are positioned frontally facing the subject and in a 30° angle.

the factors viewing perspective, performed exercise, targeted joint angle, on whether the hip moved during an exercise, and on whether a joint angle was an upper or lower body angle. The hip center movement was selected as the center of *ARKit*'s coordinate system is located at the hip center, which could lead to instabilities in *ARKit*'s tracking when moved. The upper/lower body angles were selected as laboratory experiment A indicated differences in the correlation of the upper body angles (shoulders and elbows) and lower body angles (hips and knees).

2.3 RQ 3: In which fitness, medical, or health-related use cases can optical mobile motion capture systems be applied?

Several Case Studys (CSs) were performed to investigate the possibilities and limitations of physical activity recognition, specifically optical motion capture, in mobile sports and health applications. For each CS, separate sub-questions of RQ 3 were defined. The resulting CSs contain insights into pattern recognition using *ARKit*, usability and user experience evaluations, and an outlook on collecting physical activity data and other health data for scientific studies. This section summarizes the used methods of Publications [A] (RQ 2), [C], and [D]. Further details on the used methods and results can be found in Chapters 4, 6, and 7.

2.3.1 CS 1: Goalkeeper Training

The first CS was performed in goalkeeper training (Publication [A], CS 1 [55]). As goalkeepers only represent a small part of the team, trainers often lack time to focus on their training. An automated tool for recognizing goalkeeper exercises and identifying potential mistakes during the execution is proposed to be beneficial to support goalkeeper training. CS 1 focused on the applicability of an *ARKit*-based tracking to recognize and classify goalkeeper exercises under the underlying research question:

- **RQ 3.1:** Does *ARKit* record sufficiently accurate data to identify movement-specific patterns, which can be used for movement classification?

During the CS, a dataset of 15 different goalkeeper exercises was recorded from two positions using *ARKit* (Figure 2.4). A third smartphone was used to film the exercises, which was the basis for labeling the data afterward. We included 15 exercises in the training: six dive variations, five catch variations, two throw variations, and two kick variations. Four goalkeepers were recorded. Their trainer observed the correct execution

Figure 2.4: Overview of the data recording setup used for recording the goalkeeper exercises.

of the exercises. Ten training sessions were recorded and resulted in a data set of 1050 exercise executions in total. For pattern recognition, machine learning using Convolutional Neural Networks (CNNs) was applied. CNNs proved to have a better exercise recognition performance than Long-Short-Term-Memory Networks (LSTMs) and Deep Neural Networks (DNNs) with an accuracy of 75%.

2.3.2 CS 2: Physiotherapy and Resistance Training

The second CS was performed in physiotherapy and resistance training (Publication [A], CS 2 [55]). In both sports and resistance training, patients and athletes perform specific exercises to strengthen their muscles. While this aims to improve the musculoskeletal system, wrong exercise executions can lead to injuries. The goal of CS 2 was to build a mobile application based on optical motion tracking. The mobile application should be used for the creation of new exercises on the therapist's/trainer's side and counting exercise repetitions on the patient's/sport person's side based on the research question:

- **RQ 3.2:** How can we dynamically create new exercises and count exercise repetitions?

As part of CS 2, we developed the Augmented Reality Exercise Analysis (AREA) algorithm, which enables dynamic creation and repetition counting of exercises using a pose-based approach. AREA allows users to create new exercises by splitting them into several poses and recording them once. The exercise can then be used to detect repetitions, identify correct and incorrect executions, and provide individual feedback to the users. A detailed overview of the algorithm is detailed in Figure 2.5.

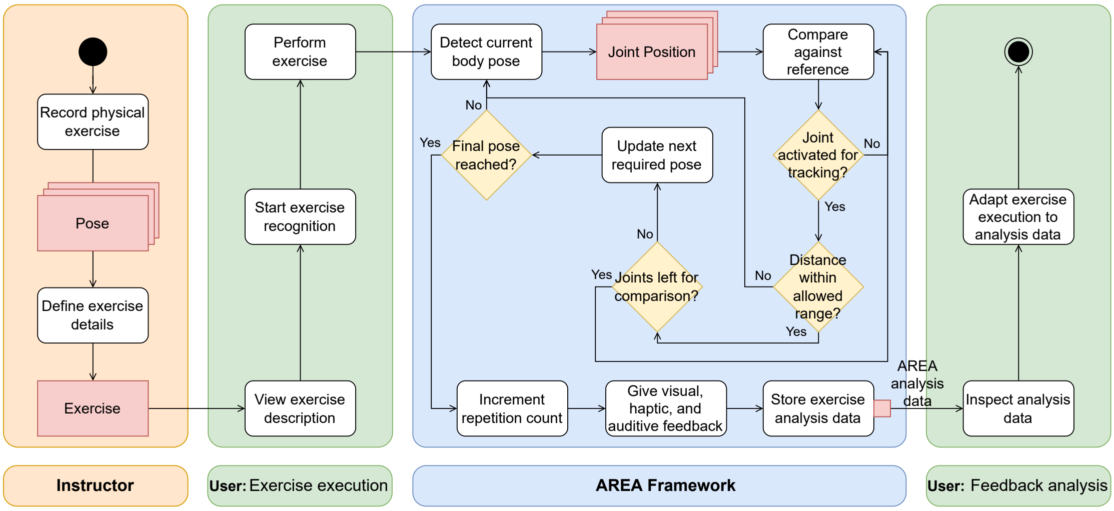


Figure 2.5: UML activity diagram representing the flow of events of the AREA algorithm. The diagram includes activities performed by the instructor to record a new exercise, the user by performing an exercise and analyzing their feedback, and the AREA framework recognizing and counting valid exercise repetitions.

2.3.3 CS 3: Golf-related Exercises

Golf is a complex sport consisting of several intra- and intermuscular movements. The golf swing requires a lot of coordination, flexibility, and strength, all of which must be practiced continuously. Even for qualified trainers, golf swing assessment and fault detection are challenging. In CS 3, we attempted to provide an automated assessment tool for a golf-related exercise consisting of three different jumps to reflect the three stages of a golf swing (Publication [A], CS 3 [55]). The jumping combination consists of a normal jump on the spot, a second jump including a 90° turn to the left or the right side, depending on the handedness of the golf player, and the third jump with maximal force. Each of the phases relates to one of the phases of the golf swing: Forward Swing, Acceleration, and Follow-Through. The most important metrics include the correct execution of the three jumps and the maximum acceleration reached while jumping. Our research questions for CS 3 were:

- **RQ 3.3:** How can we augment the data collected through optical motion capture using sensor fusion?
- **RQ 3.4:** How do qualified trainers rate *ARKit*-based applications as training support?

To answer RQ 3.3, we proposed using optical motion capture using *ARKit* to track the motion and complement the data by an additional IMU attached to the hip of the tracked person. Additionally, we performed a questionnaire with professional golf trainers on the perceived usefulness of such an application. We collected data from around 300 exercise executions with *ARKit* and the attached Inertial Measurement Unit (IMU) sensor. A professional golf trainer assigned one of five categories representing the quality to the recordings. To compensate for different execution speeds, each recording is processed using a Dynamic Time Warping (DTW) algorithm before comparison against a reference close to perfect execution (rating of 5). The measured distance to the reference execution allows the classification of the performed exercise and provides an automatic grading. A prototype of the corresponding mobile application, *Golf Coach*, was developed. We used a questionnaire to evaluate the usability and perceived usefulness of the *Golf Coach* application to answer RQ 3.4. The questionnaire was sent to 22 golf trainers. The questionnaire contained several screenshots of the application and explained the app's concept. The golf trainers were asked whether they would use the *Golf Coach* app and which features would be most important to them. Additionally, they were asked to state whether they would use apps in general to track their trainees' progress and whether they had used apps for training before.

2.3.4 CS 4: Balance Assessment

Postural imbalance is a symptom of different diseases. These include neurological diseases, instabilities of the musculoskeletal system, or conditions of the vestibular system. Regular balance tests are crucial to assess the severity, detect a progression of the imbalance early, and prevent falls. In CS 4, we investigated the feasibility of performing such an assessment by using mobile optical motion capture (Publication [C] [58]). Moreover, we investigated the usability of a mobile app for self-assessments using the smartphone camera. The research questions investigated as part of CS 4 were:

- **RQ 3.5:** How accurate can we perform a balance assessment using optical motion capture?

- **RQ 3.6:** How can we guide users to perform a balance self-assessment in front of the smartphone camera?

We implemented a prototypical mobile application, *mBalance*, which digitizes the Romberg test (Figure 2.6). The Romberg test is a validated balance assessment, indicating imbalance if the patient cannot maintain their balance in two consecutive test runs. In both runs, they have to stand in a specific test position for 60 seconds, once with open and once with closed eyes. *mBalance* observes the body joints using *ARKit* and detects whether the subject is in the Romberg test position. The test position is specified as standing with the arms crossed in front of the chest and feet next to each other. A healthy subject is supposed to be able to stay in this position without opening their arms or legs to maintain balance. *mBalance*, therefore, detects loss of balance if the joint positions of the left and right wrist indicate that the arms are not crossed anymore, or the subject has taken a step. As the subjects need to film themselves during the assessment and cannot see the smartphone's screen, they are guided through the assessment using audio commands.

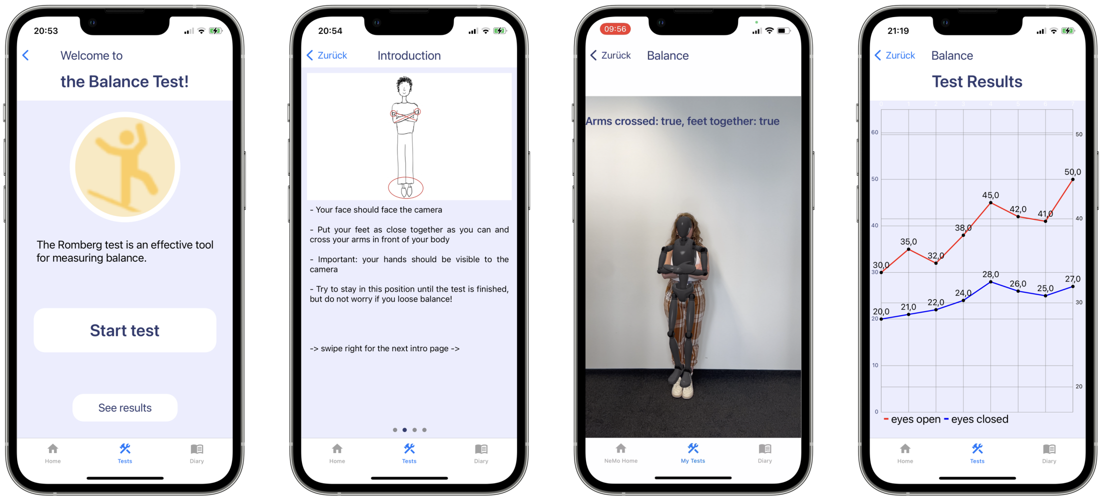


Figure 2.6: The main screens of the *mBalance* application: the start view, one of the introduction pages explaining the testing position, the test view itself, and the results overview.

To answer RQ 3.5, we performed a laboratory experiment with 31 healthy participants. All participants were asked to perform the Romberg test using the *mBalance* application three times. In the first two test runs, they were asked to maintain balance, while in the third run, they had to indicate a balance loss by taking a step or lifting their arms. The study conductor observed the participants, and the app's response was noted. The study observer indicated a true loss of balance if the participant left their test position as described above.

A second evaluation was performed to investigate RQ 3.6. 30 study participants were asked to perform one balance assessment independently and use the Thinking Aloud technique to allow insights into their thought processes while using the app. Following the Thinking Aloud protocol by Jakob Nielsen [59–61], the participants were instructed to use the app to perform a balance assessment and to say every thought they had while using the app loudly. In addition, they were told that they could use a provided tripod to position the phone. Apart from that, no further instructions were given, and no questions were answered during the session. The participants were recorded during the test sessions, and their responses were grouped into categories. These categories were quantitatively analyzed to gain insights into possible usability problems, which the study conductors predefined.

2.3.5 CS 5: Prevention of Cardiovascular Diseases and Scientific Data Collection

According to government regulations like the European General Data Protection Regulation (GDPR), health data is sensitive user data and must be protected. The same regulations apply to physical activity data as part of health data, especially in use cases like CS 4. While one of the benefits of mobile motion capture solutions like Apple *ARKit* is that they can analyze motion locally on the device, the motion data could be interesting for scientific studies and improve underlying algorithms. For this, a secure mechanism to perform scientific data collection is needed, which motivates the research question:

- **RQ 3.7:** How can we securely store sensitive user data in sports- and health contexts for scientific contexts?

RQ 3.7 was investigated as part of a mobile application to prevent Cardiovascular Diseases (CVDs) and enable scientific data collection (Publication [D] [62][1]). CVDs are the major leading cause of death worldwide [63]. Among other risk factors such as smoking, diabetes, or obesity, lack of physical activity increases CVD risk. Mobile applications have been proven to positively influence daily behavior in the short term [64, 65]. However, little data is available on the long-term impact of mobile health and fitness applications on CVD risk. Smartphones and wearables have become an essential part of

[1] Publication D is embedded for additional context and has not been considered as part of the assessment of this dissertation.

the daily life of large parts of the population and could give insights into their owners' lifestyle even years before a CVD is diagnosed. However, studies investigating these lifestyle factors require collecting sensitive health data over more extended periods of time. This demands secure storage of the data in anonymized form. The research around CS 5 proposes a mobile application serving as a mobile life coach to prevent CVDs and as a platform for data collection. While the life coach part of the app mainly serves as a means to motivate its users to regularly use the app and store their health data in it, the data collection platform aims at finding possible participants for a scientific study based on their health and lifestyle and offering them a way to participate in scientific studies.

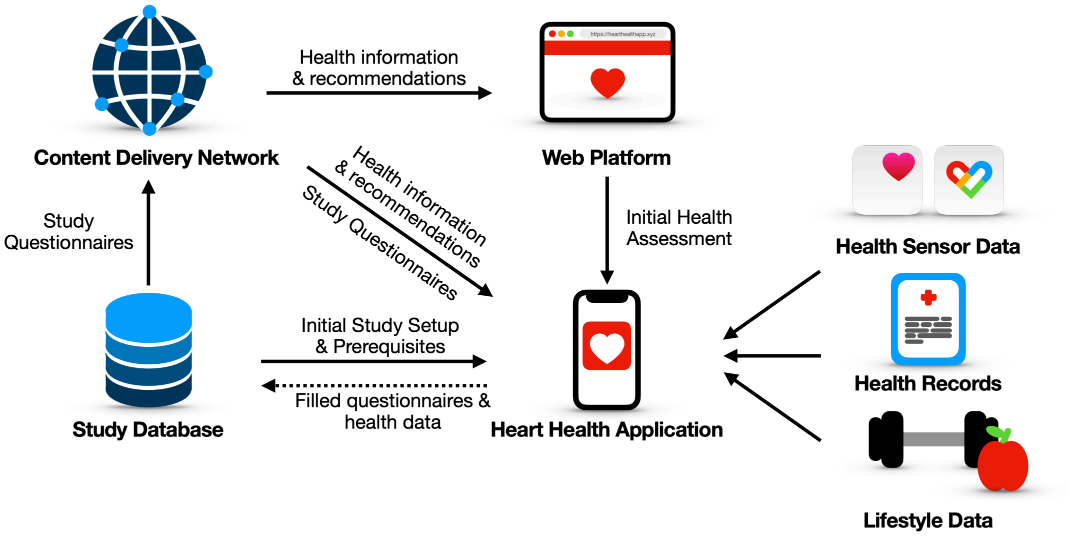


Figure 2.7: Overview of the data flow of the Heart Health Application, an example of how to perform scientific data collection on sensitive data such as health data.

Figure 2.7 explains the data flow of the application proposed in Publication [D]. The core of the system is the Heart Health Application. App users can download it directly from the app stores or enter it through the corresponding web platform. The Heart Health Application receives health information and health recommendations through articles or videos from a Content Delivery Network, which the users can use to inform themselves about different topics in the area of cardiovascular health. It collects user data from the sensors of the smart devices, such as heart rate or blood pressure and lifestyle data. Additional data can be manually entered by the user, for instance, health record data, such as results from blood screenings. Lifestyle data can include information about the nutrition but also about physical activity. All data can be manually entered into the app or recorded via additional sensors such as a smartwatch or motion capture frameworks like *ARKit*. All data is stored locally on the user's device to ensure maximum data privacy and give the users full control over their data. Via the Study Database, new scientific studies can be added. Once a new study is available, it is received by every device running the Heart Health Application. If an app user matched the prerequisites for study participation, the app asks them whether they want to participate in the study. If they agree, their data is anonymized or pseudonymized based on the study requirements, encrypted via asymmetric encryption, and sent to the study server. Additional questionnaires can be distributed to the app users via the Content Delivery Network if required by the study. The questionnaire results are also sent to the Study Database and stored securely.

3 Background and Related Work

Physical Activity Recognition is a broad area and covers several technological solutions in various fields of applications. It is a specialization of the broader term Human Activity Recognition (HAR), which focuses on detecting and classifying all kinds of human activities. HAR includes daily activities (sleeping, cooking, eating, ...), but also more specialized activities such as physical activity. A core concept to enable physical activity recognition and analysis is Human Motion Capture (HMC). Many technologies are available to capture human motion, including optical, inertial, magnetic, mechanic, and ultrasonic sensors.

As this dissertation investigates physical activity recognition using mobile devices, this chapter also describes different sensors in mobile smart devices which can be used for this purpose. It further details and compares the most common frameworks for optical motion capture available on mobile smart devices.

3.1 Human Motion Capture

HMC describes the process of capturing human motion using a computer [66]. HMC has been a growing field of interest since the 1970s and is continuously improved and empowered by more powerful processing units integrated into computers [7]. The process can be divided into four stages: initialisation of the system, tracking the human body, estimating the body's pose, and recognizing the motion [66–69]. During the initialization phase, the motion capture system prepares the recording, for example, by loading a body model to fit onto the recorded data. During the tracking phase, the human body is tracked by the system. In the pose estimation phase, specific areas of interest of the body are estimated. In the final stage of motion recognition, the current motion is calculated based on the recognized poses. Computer systems for HMC can target the whole process or only parts of the process until a specific stage, e.g., leaving out the motion recognition phase. [66]

HMC can be applied in a variety of fields. Moeslund *et al.* define three main application fields for HMC: (1) crowd surveillance, (2) motion capture as means to interact with software through specific motions and gestures, and (3) analyzing motion for diagnostics and performance in patients and athletes. Surveillance systems focus on understanding and monitoring human crowds, for example, in public transport stations. Potential use cases include counting people, analyzing traffic, or detecting congestion. More advanced use cases include behavior and activity analysis, such as violence detection or analysis of shopping behavior [71–74]. Control and interaction systems can be applied as interfaces to games or to animate virtual characters in movies [75–77]. Diagnostics and performance analysis systems aim to generate insights into body kinematics and identify possibly problematic movements, which could promote injuries or hint at specific diseases. HMC is commonly used in gait analysis [78] or sports applications [9]. Application field (1) differs from fields (2) and (3) in the number of people captured: while (1) usually focuses on analyzing multiple humans on a more superficial level, (2) and (3) capture the motion of a limited set of people, ideally only one person, on a detailed level. [70]

This dissertation focuses on use cases in the application field (3).

3.2 Human Motion Capture Systems

The underlying technology that HMC systems use for capturing human motion can be divided into optical, inertial, mechanic, magnetic, and ultrasonic systems. Apart from the used technology, they are divided by additional factors such as the number of human bodies that can be tracked simultaneously, the mobility of the tracking system, costs of acquisition and operation, and accuracy. This section delivers more detailed insights into the technologies and compares and evaluates them for suitability in the given context.

3.2.1 Optical Systems

Optical systems capture motion by using optical sensors such as cameras. Based on the used technology, they can be further divided into Optical Motion Capture Systems (OMSs), which are marker-based, and Image Processing Systems (IPSs), which are markerless.

3.2.1.1 Optoelectronic Measurements Systems

According to van der Kruk and Reijne, OMS are the gold standard systems for motion capture [9]. Indeed, they are used as reference systems for motion capture in several studies [42, 43, 79–83]. OMS require multiple optical sensors around a subject and markers placed on specific landmarks of the subject. They are further categorized into active and passive systems, depending on whether the markers send signals to the optical sensors (active system) or not (passive system) [84]. Examples for OMS are the *Vicon*

System (Vicon, Oxford, UK) [85] or the *Qualisys* motion capture systems (Qualisys AB, Göteborg, Sweden) [86]. Figure 3.1 shows a *Vicon* system recording a woman.

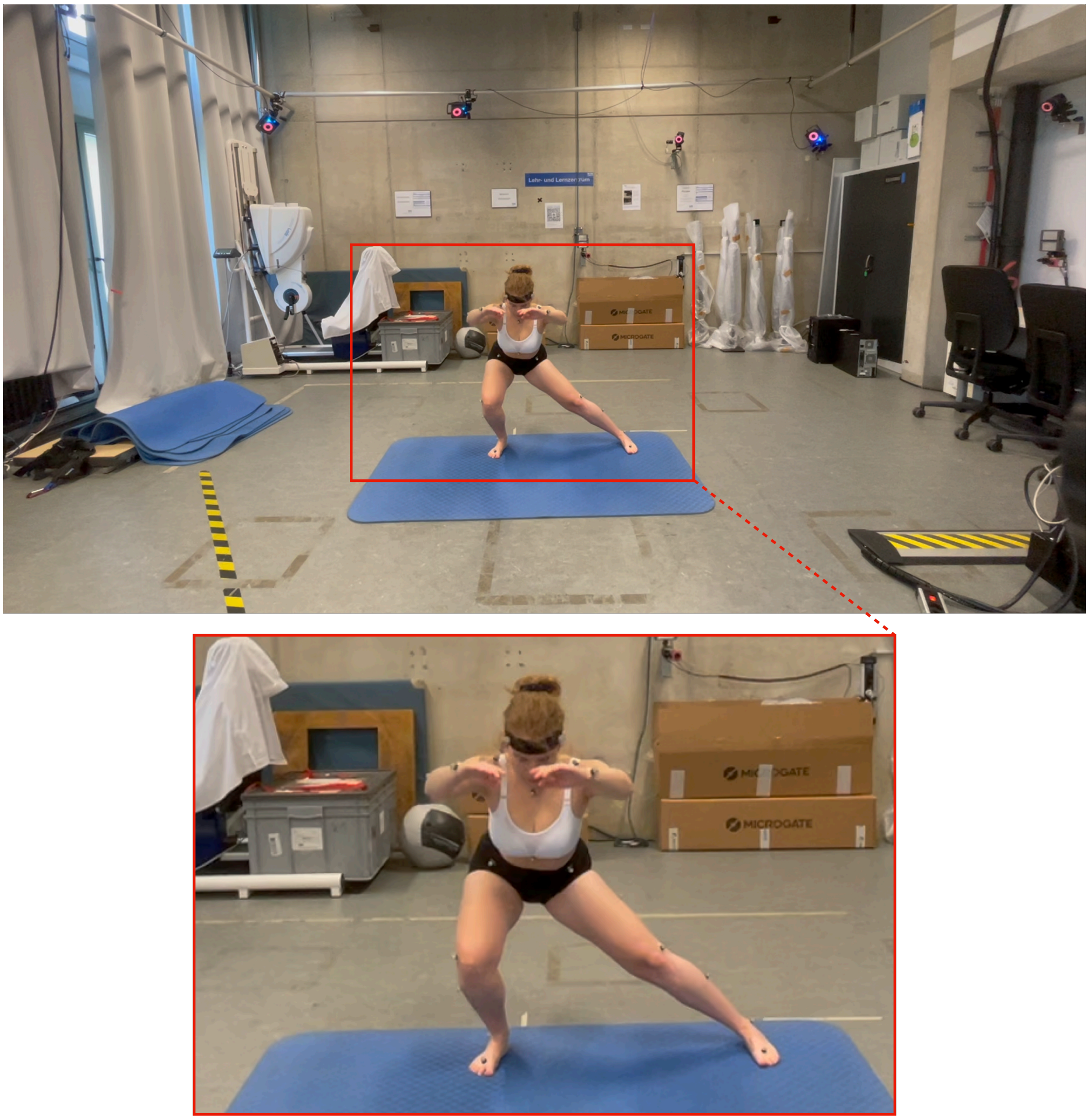


Figure 3.1: Woman being recorded by the *Vicon* system with reflective markers attached to her body. Infrared cameras are placed around the subject to capture the markers.

In their review from 2001, [66] defined several assumptions of optical motion capture, many of which are still prerequisites for good motion capture in current motion capture systems (Table 3.1). Even though OMS exhibit high accuracies, they are subject to several limitations. These limitations include the motion capture environment [87, 88], costs [89, 90], and complexity of the systems [91]. Another common problem is the

Table 3.1: Assumptions of optical motion capture systems (based on [66]).

Assumptions related to movements	Assumptions related to appearance
	Environment
1. The subject remains inside the workspace	1. Constant lighting
2. None or constant camera motion	2. Static background
3. Only one person in the workspace at the time	3. Uniform background
4. The subject faces the camera at all time	4. Known camera parameters
5. Movements parallel to the camera-plane	5. Special hardware
6. No occlusion	*Subject*
7. Slow and continuous movements	1. Known start pose
8. Only move one or a few limbs	2. Known subject
9. The motion pattern of the subject is known	3. Markers placed on the subject
10. Subject moves on a flat ground plane	4. Special colored clothes
	5. Tight-fitting clothes

limited capture volume of OMS [9], which is often limited to an area of the size of the laboratory room. While OMS are well-suited for detailed HMC, their high costs and lack of mobility limit their usage to specific scenarios.

3.2.1.2 Image Processing Systems

Advances in computer vision and mathematical algorithms enabled HMC based on image processing. IPS systems recognize motion in video data, making it applicable to various use cases, which are impossible with marker-based OMSs. IPS systems are further categorized into 2D and 3D motion capture systems, which deliver the tracked body's joint coordinates in two or three dimensions. Specific body landmarks are recognized and extracted from the video data by particular algorithms, which commonly use machine learning to accomplish the detection. Algorithms to extract body information from the Red-Green-Blue (RGB) image may use algorithms such as Convolutional Neural Networks (CNNs) or Part Affinity Fields (PAFs) [92]. Once the body is recognized, a predefined humanoid model is applied to estimate the shape and kinematic structure of the subject to track [70]. Based on this model, the recognized person's joint coordinates are determined in either two or three dimensions.

One of the most researched systems is *Kinect* (Microsoft Corp., Redmond, WA, USA). *Kinect* combines an RGB-camera with infrared sensors and captures motion in 3-dimensional space [81, 82]. Even though it is more affordable than most OMS, *Kinect* still requires a particular setup for motion capture, including the *Kinect* camera itself and a computer that it is connected to.

Recent advances in smartphones, integrated processing units, and sensors broadened the offer of IPS software, enabling HMC on smartphones and tablets. The most common

IPS software frameworks running on mobile devices are *OpenPose* (CMU, Pittsburgh, PA, USA) [93], *ARKit* (Apple Inc., Cupertino, CA, USA) [47], *Vision* (Apple Inc., Cupertino, CA, USA) [94], Tensorflow *Pose Estimate* (Google, Mountain View, CA, USA) [95], and *MediaPipe Pose* (Google, Mountain View, CA, USA) [96]. Figure 3.2 shows a recording using the *ARKit* system.

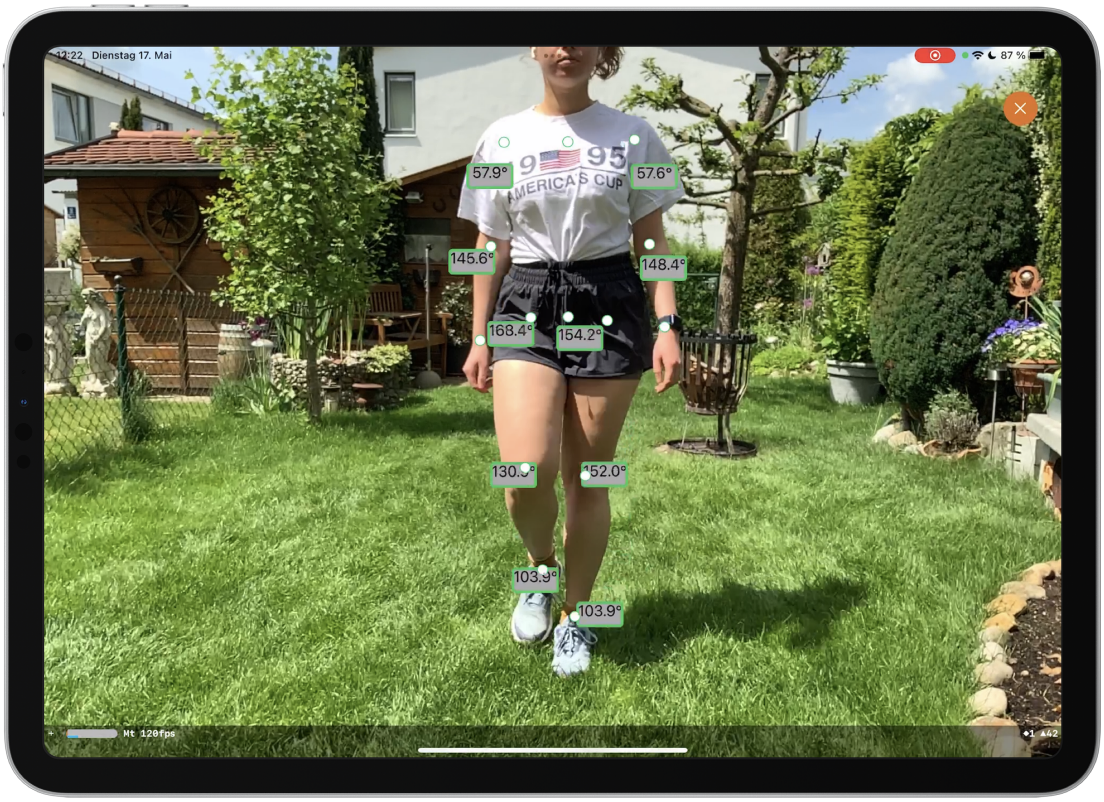


Figure 3.2: Woman being recorded by the *ARKit* system running on an iPad. The joints are detected by the camera and visualized on the screen including the detected joint angle in degrees.

IPS overcome limitations such as special hardware, all subject-related limitations except for tight clothing, slow and continuous movements, or a known motion pattern (see Table 3.1). Most importantly, they do not require the placement of markers and thus reduce the time needed for motion capture preparation and post-processing. However, other assumptions known from marker-based OMS still hold for IPS. Primarily 2D IPS still rely on the camera standing still and facing the subject from a specified angle and motion parallel to the camera plane [46]. Assumptions regarding the environment are relevant for maintaining high-quality motion recognition. The limitations are shown in several studies on 2D IPS systems. The studies provided evidence that the reliability and validity of the measurements depend on which motion is executed, the video quality, and the positioning of the camera [42–45]. Newer 3D motion capture frameworks promise to overcome these limitations. By using additional sensors such as the integrated IMUs and ambient light sensors to determine the device's position and depth sensors to gain a better scene understanding, they promise more accurate motion capture results [47–51]. IPS, particularly mobile IPS, are available on many consumer-grade devices and thus promise to be a ubiquitously available solution for making HMC accessible to a broader range of users.

3.2.2 Inertial Systems

Inertial Measurement Systems (IMSs) capture motion by using one or more Inertial Measurement Units (IMUs). IMUs contain accelerometer and gyroscope sensors to measure the IMU's motion within its six degrees of freedom. Acceleration determines the IMU's movement and position. The angular velocity can be used to calculate the IMU's rotation angle. When attached to specific parts of the body, such as joints, IMUs can capture the motion of that part. To accurately capture motion, an IMU must be attached to every region of interest of the human body, which can result in many IMUs needed to capture the whole body. Examples of IMU-based systems are the *Xsens* systems (Xsens Technologies B.V., Enschede, The Netherlands) [97] or *Perception Neuron* (Noitom Ltd., Miami, FL, USA) [98]. A woman wearing the *Perception Neuron 2* full-body suit is depicted in Figure 3.3.

The underlying assumptions for IMU-based systems differ from those for OMSs and IPSs. IMSs do not require a fixed workspace such as the OMSs and IPSs systems, nor do they need the subject to stay within the workspace or in a specific position towards the system. Also, fewer environmental limitations exist. Compared to OMSs, they track motion with lower, but still high accuracy [9, 99, 100]. Ligorio *et al.* measured less than $4°$ of Root Mean Squared Error (RMSE) [99]. Mavor *et al.* measured $8.74° \pm 1.25°$ RMSE in saggital, $5.42° \pm 1.52°$ RMSE in frontal, and $7.18° \pm 2.69°$ RMSE in transversal plane across body joints [100]. Limitations of IMSs include that the individual IMUs do not consider a fixed-space coordinate system [9] and need to be calibrated to convert the sensor orientation to the human body orientation [99]. Depending on the required part of the body to be tracked, the setup of such systems can also be complex.

3.2.3 Mechanic Systems

Mechanic Measurement Systems (MecMSs) use specialized hardware in the form of exoskeletons with integrated sensors worn by the subject to track. The sensors include

Figure 3.3: Woman wearing the Perception Neuron 2 Full-Body Motion Capture System.

goniometers and potentiometers to capture the joints' movements and transform them into electrical signals. An example of a MecMSs is the Animazoo *Gypsy 7* Motion Capture System [101]. Regazzoni *et al.* state that, while these systems are relatively affordable, they are highly invasive and can affect the motion patterns. [7]

3.2.4 Electromagnetic Systems

Electromagnetic Measurement Systems (EMSs) use the traveling time of electromagnetic waves between a transponder and base stations to detect the location of a person tagged using the transponder [102]. EMS technology is often referred to as radar. Similar to OMSs, EMSs can also be divided into marker-based and markerless systems. An example of a marker-based EMS is the Mindflux *MotionStar Wireless* [103]. Especially markerless EMSs have gained increasing attention in HMC research, as EMS sensors are available at low prices (ca. 30$). They showed highly accurate motion capture results in use cases such as hand gesture recognition or gait analysis, evaluating spatiotemporal parameters such as the knee velocity [104–106]. EMSs are able to track motion with high accuracy and low errors of $< 30mm$ [107]. Regazzoni *et al.* mention that while these systems are affordable, marker-based EMSs exhibit a narrow capturing area and wiring, which might also impact the motion. [7]

3.2.5 Ultrasonic Systems

Ultrasonic Localization Systems (ULSs) use the same principle to detect objects as EMSs but utilize the traveling time of ultrasound waves instead of electromagnetic waves to detect the location of a person [9]. ULSs require active markers attached to the targeted body joints to capture motion. They have been proposed as low-cost alternatives to established HMC systems such as OMSs or IMSs. Different studies show that ULSs deliver accurate motion data with errors of $< 55mm$ [108–110]. Qi *et al.* achieved a deviation of $< 18mm$ and $< 1.8°$ with their custom-built ULSs [111]. Ultrasonic transducers needed for building custom motion capture systems are available at 20$.

3.2.6 Comparison of HMC Systems

Each of the presented systems exposes certain advantages and disadvantages. Table 3.2 compares the motion capture technologies with regards to the price, measured accuracy in different studies, availability for consumers, whether their operation and analysis require special training, and whether a laboratory setting is required. The availability is determined by whether the products are marketed to the general public or only available to specialized institutions. The training requirement specifies whether the systems require special knowledge on operating the system, attaching markers or sensors to the body, or performing manual post-processing tasks. The requirement for a laboratory setting indicates whether a dedicated room is needed where the motion capture system is installed.

Even though marker-based OMSs deliver the highest accuracy, their high costs and extensive setup limit their usage by the general public. Similarly, MecMSs, EMSs, and ULSs are not available to the general public and require special training. While IMSs are publicly available, their interpretation for motion quality analysis still requires the knowledge of an expert. IPSs, in particular, IPS software available on smartphones and tablets is, therefore, the focus of this dissertation. A more detailed overview of mobile IPSs is given in Section 3.4.

Table 3.2: Comparison of HMC Systems.

	OMSs	IPSs	IMSs	MecMSs	EMSs	ULSs
Price	No pricing information available. Estimated costs: $> 100.000$$.	Available on most newer smartphones. Costs: 500-1000$ [112].	Full-body motion capture systems available at 3.400$ [98].	Full-body motion capture systems available at 8.000$ [101].	Markerless base stations are available at 30$.	Ultrasonic transducers are available at 20$.
Accuracy	$0.297mm$ $-6.7mm$ [9]	$< 40mm$ [113]/ $< 9.9°$[114]	$< 4°$ [99]/ $8.74° \pm 1.25°$ RMSE in saggital, $5.42° \pm 1.52°$ RMSE in frontal, $7.18° \pm 2.69°$ RMSE in transversal plane [100]	No information available.	$< 30mm$ [107]	$< 55mm$ [108]
Available for purchase by consumers	✘	✔	✔	✔	✘	✘
Training required for Operation and Analysis	✔	✘	✔	✔	✔	✔
Laboratory Setting Required	✔	✘	✘	✘	✘	✘

3.3 Sensors in Mobile Smart Devices

Different definitions exist for the terms *Smart Device* and *Mobile Device*. Both definitions are used in the Internet of Things (IoT) context. In their review from 2018, Silverio-Fernández *et al.* compare both terms. Risteska Stojkoska and Trivodaliev refer to smart devices as the core devices used in IoT [116], while Lanotte and Merro mention both terms [117] and Bisio *et al.* only refer to mobile devices in the same context [118]. Silverio-Fernández *et al.* conclude that *Mobile Devices* refer to devices with a high degree of portability. In contrast, *Smart Devices* refer to devices integrating a "certain level of embedded cleverness" [115]. However, the terms are often used interchangeably in general linguistic usage, and a fixed definition is missing. As this dissertation focuses on devices exhibiting both characteristics, this section will use the term *Mobile Smart Device* further to define the field, detail possible application areas, and contextualize it into the scope of the research.

Examples of mobile smart devices are smartphones, tablets, and wearables such as smartwatches. They usually integrate several sensors which can be used for human activity recognition and physical activity recognition in particular. Sousa *et al.* cluster the available sensors into "inertial (e.g. accelerometers and gyroscopes), magnetic (e.g. magnetometers), acoustic (e.g. temperature, barometer, humidity, gravity and pressure), environment (e.g. light and proximity), location (e.g. GPS and RFID), radio (e.g. GSM, WiFi and Bluetooth), physiological (e.g. heart beat) and software sensors (e.g. services and applications)" [119].

3.3.1 Inertial Sensors in Mobile Smart Devices

Research on HAR using mobile devices mainly uses the integrated IMUs for detecting motion [119–122]. Ronao and Cho achieved high accuracies in classifying physical activity using inertial sensor data and convolutional neural networks [120]. Mukherjee *et al.* propose using the ensemble of three methods (CNN-Net, Encoded-Net, and CNN-Long-Short-Term-Memory Network (LSTM)) to classify human activity data gathered from inertial sensors [122].

A common issue with inertial sensors is the device's orientation and position during activity recognition [121]. Different device orientations during similar activities can lead to different patterns in the recorded signals, thus complicating the analysis [119]. Sousa *et al.* evaluated different activity features. They concluded that orientation-independent features such as the device's magnitude are not suited for activity classification and that accelerometer data is sufficient to classify activities [119].

Ishii *et al.* found that wearable inertial sensors can count exercise repetitions [40]. Zelman *et al.* showed that exercise repetition counting is possible even without prior exercise-specific training of the classifier [41]. Jenny and Dietz *et al.* investigated the suitability of measuring knee motion using the inertial sensors of a smartphone. Both publications state accurate results for the approach [32, 33].

3.3.2 Optical Sensors in Mobile Smart Devices

Smartphones' and tablets' optical sensors have been the subject of increased interest for physical activity recognition. Optical sensors in mobile smart devices mainly refer to integrated active RGB cameras.

An example of motion capture on mobile using an active RGB camera is the *OpenPose* software, which was already mentioned in Section 3.2.1.2. *OpenPose* can be used on computers and mobile devices and achieves good accuracy. Next to active RGB cameras, several mobile devices include Red-Green-Blue-Depth (RGB-D) cameras. This technology allows measuring the distance between the camera and objects such as human bodies in the scene by creating a depth map [7, 123].

A commonly known RGB-D sensor is the Microsoft *Kinect* system, which has been evaluated for HMC in a variety of studies [81, 82]. However, it is designed as a stationary device and is thus not categorized as a mobile device in the context of this research.

Recent smart devices contain RGB-D sensors such as structured light sensors, usually integrated for authentication using the face of their owners, and Light Detection And Ranging (LiDAR). New software frameworks developed explicitly for smart devices combine RGB cameras for motion capture with additional sensors. They use the smart device's IMU sensors to identify the device's position and orientation, ambient light sensors to estimate its surroundings' light setting, and can include RGB-D cameras such as LiDAR or structured light sensors to generate a depth map of the captured scene [47]. First studies indicate good motion capture results of the lower limbs of a single subject [124]. More extensive validation studies of such mobile frameworks are missing. These studies should include multiple subjects and track the whole body in different movements.

3.3.3 Other Sensors in Mobile Smart Devices

While other commonly integrated sensors like light, proximity, barometer, and Global Positioning System (GPS) sensors can support HAR, their applicability for more detailed motion analysis such as exercise repetition counting or measuring knee motion is limited. The same applies to sensors commonly integrated into wearables, for example, heart rate trackers. They allow identifying that an activity is performed due to an increased heart rate but are limited to concluding the intensity of the activity.

Mobile smart devices contain various sensors that can be used to track human activity and physical activity in particular. Especially inertial and optical sensors are well-suited for tasks such as activity classification, repetition counting, and measuring motion quality. However, optical sensors allow more detailed human body tracking simultaneously than inertial sensors, which is why this research focuses on optical motion capture.

3.4 Frameworks for Optical Motion Capture on Mobile Smart Devices

Several software frameworks for markerless motion capture using active RGB cameras have been developed, as previously mentioned. They can be differentiated by whether they capture motion in 2D or 3D, the underlying algorithm for recognizing the human body and extracting the joint information from the camera data, and whether they use additional sensor data such as inertial or RGB-D sensors to enhance the tracking. In the following, we provide an overview of the most common frameworks available on mobile smart devices.

3.4.1 2D Frameworks

2D motion capture frameworks are characterized by their output for tracking human motion, projecting the motion onto a 2-dimensional coordinate system. The origin of the 2D coordinate system is either set to a fixed rectangle, e.g., the RGB image, or dynamically determined by creating a rectangle around the recognized human in each frame. Therefore, different camera positions yield different joint coordinates if only one camera is used for recording caused by the different perspectives.

3.4.1.1 OpenPose

OpenPose is among the best-known 2D motion capture frameworks which solely rely on RGB images from a single camera for motion capture. It was developed as an open-source project by Carnegie Mellon University (Pittsburgh, PA, USA) and is therefore accessible by every software developer for use and further development. It offers functionality to track several key points of the whole body (15, 18, or 21 key points), hands ($2 * 21$ key points), and the face (70 key points) of one or more people present in the image using PAFs. Only when a multi-camera setup is used, it can detect a person in 3D. A multi-camera setup contradicts the goal of an accessible and affordable setup, ideally, by using devices that a user already owns. Thus, we consider *OpenPose* as 2D framework. *OpenPose* has been ported to all common platforms, including web and mobile applications. [92, 93, 125–127]

OpenPose has been evaluated in different studies, most of which used a multi-camera setup to increase the tracking accuracy of the system. Zago *et al.* performed gait analysis using *OpenPose* and found deviations to a gold standard system of $20mm$ in 3-dimensional tracking trajectories, $30ms$ in step and swing phase duration, and $1.23cm$ in step length [45]. D'Antonio *et al.* also performed gait analysis and found that, while the trajectories were tracked accurately, the minimum and maximum joint angles were overestimated by up to $9.9°$ [114]. Zago *et al.* additionally found that gait tracking worked most accurately when the cameras were placed at a distance of $1.8m$, and one camera was positioned perpendicular to the gait direction [45]. Nakano *et al.* evaluated gait and more rapid movements such as counter-movement jumps and ball throwing. They found that even though some of the detected joint coordinates deviated for more than $40mm$ from the reference measurement, measurements deviated by less than $30mm$ in 80% of all measurements [113].

3.4.1.2 TensorFlow Pose Estimate

The TensorFlow framework by Google Inc. (Google Inc., Mountain View, CA, USA) contains various sub-frameworks for machine-learning-related tasks and libraries for real-time pose estimation. TensorFlow *Pose Estimate* includes *MoveNet*, the most recent implementation of its motion capture algorithm, running on iOS and Android. The *Pose Estimate* frameworks use CNNs for key point detection and grouping them into poses. The resulting body model contains 17 key points. Moreover, they allow tracking multiple persons at once. [95, 128, 129]

While some researchers used the TensorFlow *Pose Estimate* framework in their applications [130, 131], no validation studies are available on its accuracy compared to reference systems.

3.4.1.3 Apple Vision

Similar to the TensorFlow framework, Apple *Vision* contains several frameworks for vision-related machine learning tasks and also includes functionality for human pose estimation in 2D. Apple *Vision* is available on iOS and macOS platforms and developed by Apple Inc. (Apple Inc. Cupertino, CA, USA). It detects 19 key points of the human body and includes real-time functionality for hand tracking (21 key points). Most recently, *Vision* has been complemented by functionality to automatically detect and count executions of repeating movements [132]. Apple *Vision* is proprietary software, and details about the implementation are not publicly documented. [94]

Up to this point, no published research projects used the *Vision* library, so no preliminary information on its accuracy is available.

3.4.2 3D Frameworks

In contrast to 2D frameworks, 3D frameworks are characterized by their ability to detect human poses in 3-dimensional space with a single device. This can be achieved by estimating the depth of the scene and the device's movement through additional sensors such as inertial or depth sensors or multi-focus image fusion [133]. The 3D tracking promises to be more independent of the capturing angle than 2D frameworks, which is why they are investigated in the scope of this dissertation.

3.4.2.1 Apple ARKit Body Tracking

ARKit Body Tracking is a 3D motion capture framework integrated into Apple Inc.'s *ARKit* framework for Augmented Reality. It tracks 93 different key points and distinguishes actively recognized key points and calculated key points, which are calculated based on the positions of the neighboring joints. Keypoints mainly represent body joints but can also define the location of other important landmarks, such as the eyes and ears of a person. Calculated joints include smaller key points that are not easily recognizable, such as toe and finger joints. Like Apple *Vision*, *ARKit* is proprietary software. Its implementation is thus not publicly documented. *ARKit* only provides live tracking and cannot be applied to pre-recorded videos. It is further limited by only using the smartphone's rear camera and can only be used on iOS devices with an A12 processor or newer. It is thus only available to a limited set of users. However, *ARKit* leverages a variety of additional sensors for motion tracking, which are not supported by most other frameworks. *ARKit* uses the smartphone's inertial sensors to detect its position in 3D space and creates a depth map through a LiDAR sensor in newer devices. With the dedicated capabilities of the M1 processor [134] and the neural engine that performs machine-learning tasks, this promises good motion capture accuracy. Therefore, this dissertation focuses on evaluating and validating *ARKit* for HMC and physical activity recognition. [47]

Initial insights into the accuracy were given by a study by Tran et al., where they measured an average deviation from the reference measurement by 2° [124]. However, more detailed studies were missing and conducted as part of this research.

3.4.2.2 MediaPipe Pose/BlazePose

MediaPipe Pose/BlazePose is a 3D motion capture framework by Google Inc. It detects 33 key points from RGB images and videos using a supervised heatmap, offset, and regression approach. In Bazarevsky *et al.*'s comparison against *OpenPose*, *MediaPipe Pose* performed significantly worse on an extensive motion data set but outperformed *OpenPose* when applied to a Yoga data set. Moreover, it showed a significantly increased performance concerning processing time. [48]

No additional evaluations of *MediaPipe Pose* exist, as it is a relatively new framework. As it does not use additional sensors for depth estimation and device orientation detection, which are supposed to improve motion capture performance, it will not be the focus of this dissertation.

4 Publication A: Mobile Motion Tracking for Disease Prevention and Rehabilitation Using Apple ARKit

This publication delivers insights into the possibilities of mobile motion tracking using Apple *ARKit* in disease prevention and rehabilitation and presents possible use cases for such applications.

Conference: dHealth 2021 - Navigating healthcare through challenging times
Number of Pages: 9
Type: Full Paper
Review: Peer Reviewed (2 Reviewers)

Summary

Problem. Physical activity is crucial to living a healthy life and helps improve the overall quality of life. Performing a physical exercise correctly is essential for performance in sports and helps prevent injuries and diseases. Rehabilitation and physiotherapy exercises help rebuild strength and relieve pain when correctly performed. Training and feedback are necessary to learn the correct execution of exercises. However, only a few solutions for automated analysis and feedback exist, which are widely accessible and can be applied to performance improvement in sports, disease prevention, and rehabilitation. **Objective.** This research presents an initial validation of the Apple *ARKit* framework for optical mobile motion capture against the *Vicon* system, a gold standard system for optical motion capture. Through *ARKit* or similar technologies, new possibilities for automated motion analysis using consumer-grade devices are enabled. Moreover, it provides insights into three case studies where such a framework could be applied in the sports and health sector: (Case Study (CS) 1) goalkeeper training, (CS 2) physiotherapy and resistance training, and (CS 3) golf. **Methods.** We performed a laboratory experiment with 12 subjects, where we evaluated the performance of mobile motion capture using *ARKit* in different exercises. The exercises included daily activ-

ities such as walking and running, catching a ball, and body-weight exercises such as Jumping Jacks, Cuttings, and Squats. The analysis compares the Euler Angles of the shoulder, elbow, neck, knee, and ankle angles calculated by *Vicon* and *ARKit*. Based on the Euler Angles, the minimum, maximum, mean, standard deviation, range of motion, and Pearson correlation were calculated for all exercises and used to compare *ARKit*'s accuracy against the reference. The three case studies present insights into how the *ARKit* data can be used for exercise classification, exercise counting, and sensor fusion with Inertial Measurement Unit (IMU) sensors to improve the tracking accuracy. In addition, a usability and usefulness questionnaire about the golf use case was evaluated. **Results.** The results of the *ARKit* evaluation exposed several limitations in tracking the full range of motion. However, the good correlation results showed an accurate tracking of the motion course. CS 1 proves that the *ARKit* data is sufficiently accurate to distinguish between exercises. CS 2 presents a pose-based approach, which can be applied for exercise repetition counting in physiotherapy and resistance training. The approach allows the dynamic creation of new exercises without the extensive data collection required by many machine learning algorithms. In CS 3, we could show that 95.5% of the questionnaire participants would use an *ARKit*-based app for their training. **Conclusion.** While *ARKit* showed limitations in tracking the range of motion, it showed good accuracy in tracking the course of motion. In three different case studies, we were able to show that (1) the *ARKit* data provides sufficient accuracy to classify exercises, (2) algorithms can be developed to enable the dynamic creation of exercises and perform exercise repetition counting without extensive data collection, and (3) *ARKit*-based golf training applications are considered highly useful by golf trainers.

Contributions

The author of this dissertation is the main author of this contribution and contributed substantially to the conceptualization of the research, the writing of the paper, the implementations of the used applications, the conduction of the case studies, and the visualizations. She conceptualized the paper and its content, including the laboratory experiment and the three case studies. Three students performed the laboratory experiment and CS 3 as part of thesis projects. The author of this dissertation supervised the theses and summarized the results within this publication. She planned and conducted case studies 1 and 2, including the implementation of the data recording application, the data recordings and analysis (CS 1), and the development of the algorithm (CS 2). Furthermore, she performed the literature review and visualized the contributions presented within this publication.

Navigating Healthcare Through Challenging Times
D. Hayn et al. (Eds.)

doi:10.3233/SHTI210092

Mobile Motion Tracking for Disease Prevention and Rehabilitation Using Apple ARKit

Lara Marie REIMER[a,1], Severin WEIGEL[a], Florian EHRENSTORFER[a], Malintha ADIKARI[a], Wolfgang BIRKLE and Stephan JONAS[a]
[a] *Technical University of Munich, Munich, Germany*

Abstract. Background: Physical activity helps improve the overall quality of life. The correct execution of physical activity is crucial both in sports as well as disease prevention and rehabilitation. Little to no automated commodity solutions for automated analysis and feedback exist. Objectives: Validation of the Apple ARKit framework as a solution for automatic body tracking in daily physical exercises using the smartphones' built-in camera. Methods: We deliver insights into ARKit's body tracking accuracy through a lab experiment against the VICON system as Gold Standard. We provide further insights through case studies using apps built on ARKit. Results: ARKit exposes significant limitations in tracking the full range of motion in joints but accurately tracks the movement itself. Case studies show that applying it to measure the quantity of execution of exercises is possible. Conclusion: ARKit is a light-weight commodity solution for quantitative assessment of physical activity. Its limitations and possibilities in qualitative assessment need to be investigated further.

Keywords. mHealth, Mobile Applications, Fitness Trackers, Augmented Reality

1. Introduction

Regular physical exercise is known to be beneficial for the overall quality of life. It improves functional capacity and reduces long-term risks for diseases like Diabetes mellitus Type 2 and Alzheimer's while improving overall health, health-related aspects, and mood [1, 2, 3]. Physical exercise can also positively influence hemic and oncological diseases' treatment and rehabilitation by preventing muscular atrophy and improving the patients' mood [4, 5, 6]. However, the correct execution of physical exercise is essential. Wrong exercise execution can lead to biochemical stress, injuries, and osteoarthritis in the respective joints [7]. To avoid wrong movements, regular supervision of the exercising person by experienced personnel is crucial. Various systems have been developed and evaluated to allow a more profound analysis of human motion and detect problematic movements in exercises. These systems include optical, magnetic, inertial, or mechanic sensors to detect and measure different kinds of metrics about human body motion [8, 9]. Modern motion capture systems track indoor and outdoor activities with different accuracy based on the underlying technologies [10]. However, none of those

[1] Corresponding Author: Lara Marie Reimer, Technical University of Munich, Germany, E-Mail: reimer@tum.de

systems currently allows daily usage as they require custom hardware for tracking and often are connected to high costs.

Many people track their daily lives with mobile devices like smartphones, -watches, and sleep or fitness trackers [11]. These devices are equipped with more advanced sensors such as camera systems, gyroscopes, accelerometers, and optical sensors. Also, the devices themselves are getting more powerful and allow resource-intensive applications, e.g., for on-device machine learning. These advances include powerful software applications to perform automatic detection of objects, among them the human body. Software frameworks for mobile devices like Apple ARKit[2] or Vision[3] or Google's Tensorflow Pose Estimation[4] automatically detect the joints' positions through image recognition techniques. These positions can be used for further analysis and enable various new applications, e.g., in mobile games and healthcare.

Using motion detection applications on mobile devices could add to the user's health, allowing quantitative analysis of physical activity. The application area is broad and includes prevention of injuries and sickness, progress, and aggravation tracking, e.g., in rehabilitation, during treatment, or as a motivation to be more active and exercise in general [11]. Mobile devices might be able to deliver a low-cost, commodity alternative to established solutions. Boulos et al. state that while many mobile applications for tracking physical activity have been developed, most of them rely on few sensor data. Used data is often limited to GPS and heart rate. Especially in the area of resistance exercises, personalized coaching applications are missing [12]. However, systems based on wearable sensors can recognize and count exercises [13], even without prior, exercise-specific training [14]. Current research shows that approaches using Apple ARKit can track the lower extremities' motion [15]. A more detailed evaluation of the applicability of ARKit, including the upper extremity, is missing. The aim of this work is to evaluate the suitability and applicability of on-device motion tracking using Apple ARKit and give an outlook to their application in three different use cases: goalkeeper training, physiotherapy and resistance training, and golf.

In all three use cases, complex motion needs to be tracked. If executed incorrectly, these exercises can overstress several joints and promote injuries. The usage of mobile motion capture to assess motion and detect incorrect movements could balance training, prevent injuries, and provide progress supervision. Existing approaches in goalkeeper training build on custom sensors to analyze motion [16]. In our approach, we rely on ARKit as an alternative approach using a commodity device (Case Study 1). Resistance training plays a substantial role in improving and maintaining physical strength and fitness, either during the prevention of diseases and injuries or physiotherapy. Research proves that active physical exercise is among the crucial factors in physiotherapy of common conditions, e.g., in chronic low back pain [17]. We design an algorithm to capture and assess body-weight exercises and provide individual feedback in a mobile coaching application in resistance training (Case Study 2). Golf consists of complex motion sequences (intra- and intermuscular) that require a high amount of coordination, flexibility, and strength, which needs to be practiced continuously to generate high performance. Even for professional golf trainers, an individual assessment of the golf swing and related exercises and tracking progress is challenging. Several solutions have been developed to support golf training. Existing solutions use different sensors attached

---

[2] https://developer.apple.com/augmented-reality/arkit/
[3] https://developer.apple.com/documentation/vision
[4] https://www.tensorflow.org/lite/models/pose_estimation/overview

to the body and the golf club to measure motion [18, 19] or focus on camera-based techniques, like the Coach's Eye application[5], which allows calculating joint angles in a single, selected video frame manually. We propose a more light-weight, automated approach by combining an optical sensor and a single Inertial Measurement Unit (IMU) sensor for qualitatively assessing motion in golf-related exercises (Case Study 3).

2. Methods

Within the scope of this paper, we aim at investigating two research questions:

- **RQ 1**: Which accuracy does Apple ARKit provide in contrast to the VICON system?
- **RQ 2**: What are potential use cases for an ARKit-based system?

To answer **RQ 1**, we performed a lab experiment in which we compared motion data generated by Apple ARKit against the VICON system's motion data. To answer **RQ 2**, we are conducting several case studies in different health and exercise science areas, each of them consisting of a mobile prototype application running the ARKit framework.

2.1. Suitability of ARKit for Motion Recognition and Tracking

A total of 12 subjects participated in the experiment, 5 females and 7 males, all of good health without physical impairments. The participants' height ranged from 1.56m to 1.96m and their weight from 52.2kg to 97.5kg. In the lab experiment, subjects had to perform 9 different exercises focusing on both the upper and lower extremities, including running on a treadmill at 3 different speeds, passing and catching a ball, jumping jacks, cuttings, and squats. All joint angles are calculated using Euler angles in 3 dimensions: x referring to flexion/extension, y referring to inversion/eversion, and z referring to the rotation. Both systems measured shoulder, elbow, neck, knee, and ankle angles. We calculated minimum, maximum, mean, standard deviation (SD), and range of motion (ROM) and compared the ARKit values against the VICON angles for all exercises.

For the study setup, we used a 10-camera VICON setup with the full-body Plug-in Gait model provided by VICON Nexus[6]. For the recording with ARKit, we placed an iPad Pro 11" 2020 with LiDAR sensor on a tripod in front of the subject with a distance of 3m.

2.2. Case Study 1: Recognition of Exercises in Goalkeeper Training

In our first case study, we wanted to identify patterns specific for individual exercises, which is the baseline for further analysis of the motion. For this, we recorded 15 different exercises with ARKit and an additional camera, which served as the basis for labeling the data afterward. The exercises consisted of 6 dive variations, 5 catch variations, 2 throw variations, and 2 kick variations. 4 goalkeepers were recorded. Their football trainer was supervising the recording sessions to guarantee the correct execution of the exercises. Through 10 session recordings in total, we created a dataset of 1050 single exercise executions mapped on the 15 exercises. We used machine learning techniques to train a classifier predicting the matching. We reduced the 6 different dive classes to 2

---

[5] https://www.coachseye.com
[6] https://docs.vicon.com/display/Nexus26/Full+body+modeling+with+Plug- in+Gait

classes due to similarities in the execution as a preparation. Based on research by Ronao et al. [20], we applied an approach based on Convolutional Neural Networks (CNN).

2.3. Case Study 2: Recognition of Body-Weight Exercises in Physiotherapy and Resistance Training using a Pose-based Approach

In case study 2, we developed an algorithm to recognize and assess body-weight exercises using ARKit. The algorithm allows app users to record new exercises, store them in a database, detect correct exercise executions of stored exercises, detect incorrect repetitions, and provide individual feedback to the user on improving the motion. Due to the COVID-19 pandemic, an evaluation of the prototype application remains open.

2.4. Case Study 3: Classification of Golf-Related Exercises

As part of case study 3, we developed a system consisting of an ARKit-based mobile application prototype and an IMU sensor to measure motion, rotation, and acceleration in an exercise related to the golf swing. The camera is placed in front of the player. The IMU sensor is attached to the hip. The system automatically assesses the exercise execution compared to a reference recording. The automatic assessment is performed by a Dynamic Time Warping (DTW) algorithm, which compares the exercise data against the reference data and evaluates the distance between the compared time series. Through the DTW approach, we classify the executions and provide a grading. We recorded around 300 exercise executions, and a golf trainer assessed them to generate training data. A prototype was implemented, but the automatic assessment could not be qualitatively validated due to the ongoing COVID-19 pandemic. Through an online questionnaire, we gained insights into how golf trainers assessed the prototype's usefulness.

3. Results

3.1. Suitability of ARKit for Motion Recognition and Tracking

The comparison of the Euler angles calculated by VICON against the Euler angles calculated based on the ARKit data exposed several deviations between the two systems across all tracked joints (see Table 1 for one joint). Across all joints and dimensions, the ARKit measurements revealed smaller SD and ROM values than the VICON measurements. Additionally, ARKit and VICON use different reference points to calculate the Euler angles, which led to mirrored values in the analysis (Figure 1b).

Table 1. Results of the VICON and ARKit measurements in the squat exercise. The values represent the knee joint angles. All values are calculated in degrees.

System	Angle	Dimension	Min	Max	Mean	SD	ROM
VICON	Knee	x	0.57	119.61	46.35	42.45	119.04
	Right	y	-3.60	16.52	5.86	6.73	20.12
		z	-22.86	36.30	1.53	20.97	59.16
	Knee	x	0.00	117.98	46.58	42.22	117.97
	Left	y	-3.50	13.77	4.46	5.76	17.27
		z	-17.14	41.08	7.69	20.29	58.22
ARKit	Knee	x	23.15	54.37	29.44	6.39	31.21
	Right	y	-9.07	1.28	-2.81	2.13	10.34
		z	-2.54	18.12	4.66	4.45	20.66
	Knee	x	22.77	55.88	28.79	7.11	33.11
	Left	y	-2.08	5.79	1.66	1.48	7.87
		z	-19.85	3.00	-5.20	4.31	22.85

Even though the results exhibited a considerably smaller range of motion, the motion itself was detected in every repetition of the exercises (Figure 1). The ROM was consistently smaller throughout all repetitions in all exercises, supported by the smaller

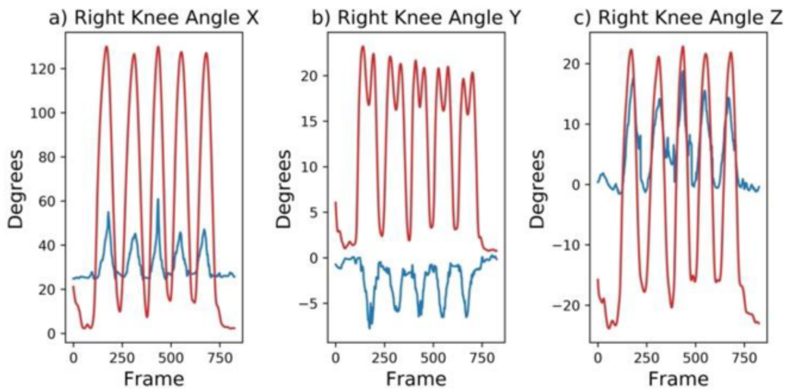


Figure 1. Knee joint angles during squat by ARKit (blue) and VICON (red).

SD values. We calculated the Pearson correlation-coefficients for shoulder, elbow, knee, and ankle angle for all participants for the squat exercise for further analysis. We calculated the mean and SD values across all participants, dimensions, and left and right sides based on the coefficients. We used the correlation coefficients' absolute values for the correct detection of the motion, as due to the different reference points of VICON and ARKit, the values of the y dimension of ARKit were mirrored for the joints on the right side of the body. The mean correlation-coefficients and SD for the shoulder, elbow, knee, and ankle angles were 0.406 ± 0.236, 0.085 ± 0.117, 0.705 ± 0.144, and 0.654 ± 0.184, respectively.

Table 2. Pearson correlation-coefficients of the VICON and ARKit measurements for all 12 participants in the Squat exercise.

	Shoulder		Elbow		Knee		Ankle	
	L	R	L	R	L	R	L	R
Min	0.017	0.011	0.000	0.000	0.297	0.410	0.222	0.207
Max	0.801	0.825	0.658	0.447	0.901	0.898	0.909	0.911
Mean	0.367	0.439	0.084	0.085	0.674	0.736	0.614	0.693
SD	0.227	0.241	0.128	0.104	0.147	0.134	0.173	0.186

3.2. Case Study 1: Recognition of Exercises in Goalkeeper Training

Figure 2 visualizes the output of the 3-dimensional positional data of a session with recordings of 5 exercises for the right hand and right foot. The different patterns are identifiable, as well as the repetitions of each exercise.

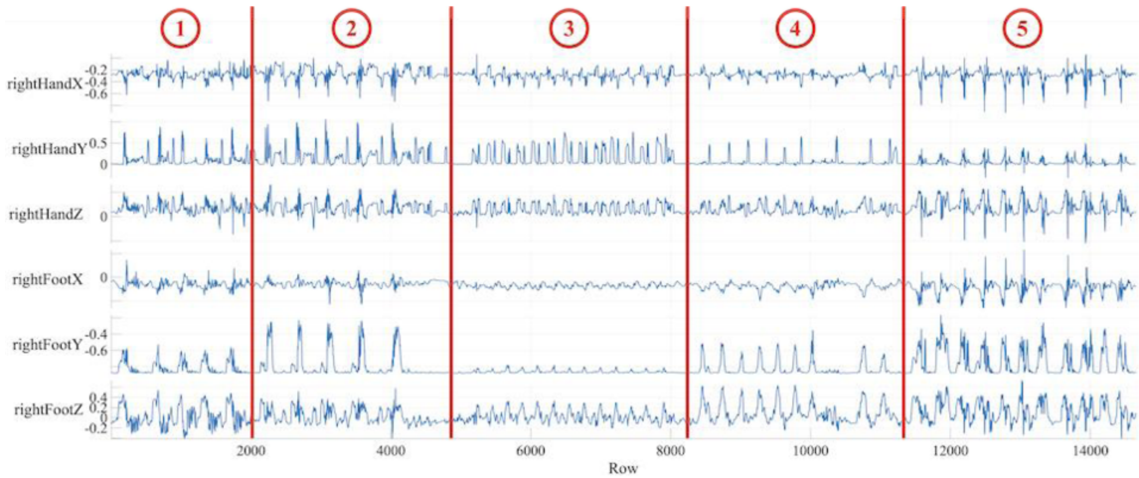


Figure 2. Dataset of a session including 5 exercises: Dive High Right (1), Dive High Left (2), Catch Hand (3), Catch Body (4), and Catch Ground (5).

Despite the comparably small dataset, the classifier achieved a validation accuracy of around 75% using the CNN in the test set (Figure 3). The dives and the two versions of the jump catch achieved high rates of correct predictions with around 70-80%. For comparison, we tested the same approach with Long Short-Term Memory Networks (LSTM) and Deep Neural Networks (DNN). Due to the small dataset, both alternatives

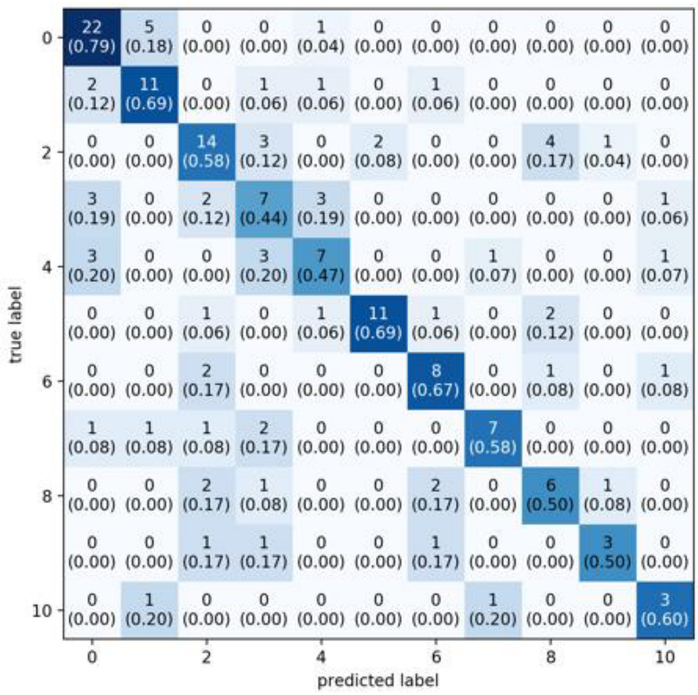


Figure 3. Confusion matrix with precision values of the CNN prediction on 11 classes: (0) Dive Right (1) Dive Left (2) Catch Hand (3) Catch Body (4) Catch Ground (5) Jump Catch (6) Jump Catch Run (7) Throw Low (8) Throw High (9) Side Kick (10) Side Kick Ground.

showed overfitting tendencies and did not reach the accuracy of the CNN with an accuracy of 55% and 65%, respectively, for all classes.

The first case study shows that the data generated by ARKit is accurate enough to distinguish between goalkeeper exercises. Even with a small dataset, a machine learning classifier can classify around 75% of the exercises correctly.

3.3. Case Study 2: Recognition of Body-Weight Exercises in Physiotherapy and Resistance Training using a Pose-based Approach

To build a mobile, ARKit-based application enabling tracking various dynamic body-weight exercises, we designed a workflow for app users to add new exercises to the app. The prerequisite for registering a new exercise is that it has to consist of clearly different poses. An exercise needs to expose motion in at least one trackable joint in three poses as observable in a squat with full extension of both knees, flexion of at least 90deg in both knees, and a full extension of both knees again (Figure 4). The system captures the three poses, including the measured angles. For each pose, a set of essential joints can be defined, e.g., knee, ankle, and hip joint in the squat. This setup should be done by qualified personnel, e.g., a physiotherapist, to ensure the exercise's correctness.

Figure 4. Exercise repetition tracking using the pose-based approach. ARKit's recognition is shown by the robot overlay. ARKit tracks the poses in a, b, and c, and updates the repetition count after pose c is matched.

For exercise recognition, we convert the 3-dimensional positional joint data provided by ARKit into angles. Our proposed algorithm observes every motion detected by ARKit and matches the motion's progression against the curve expected by the exercise. Once all poses have been reached in the correct order, the algorithm considers the repetition as completed and increases the count, as shown in Figure 4c. The algorithm is flexible enough to recognize a variety of body-weight exercises without training a specific machine learning model through this approach.

Considering the initial lab experiment results, approaches on qualitatively assessing exercise repetitions need to be investigated. Using the proposed algorithm allows the creation and tracking of dynamic body-weight exercises based on joint angle calculations.

3.4. Case Study 3: Classification of Golf-Related Exercises

We evaluated the Golf Coach app's usability in an online questionnaire. In total, 22 golf trainers. The trainers' age ranged between 19 to 66 years, with a mean of 38.5 years and a median of 32 years. Their experience as a golf trainer ranged from 1 year to 40 years, with a mean of 13.14 years and a median of 7.5 years. 95.5% stated that they would use an app to track their trainees' progress, and 70% had used mobile coaching apps before. After a guided tour through the app, the participants were asked whether they would use it. 59% stated that they would use it, 9% would use it from time to time. 9% would prefer to test it first. 5% said that they would not use it. 18% chose not to answer. 55% stated that progress tracking would be the most important feature to them.

4. Discussion & Future Work

The lab experiment and the case studies provide evidence that mobile applications based on ARKit can track joint motion. Table 1 shows that the recognition is not as accurate as the reference values detected by VICON. ARKit does not seem to capture a full extension of the knee, as the minimum value provided is 22.77deg, compared to 0.00deg seen by VICON. Flexion is recognized, but to a much smaller extent than by the VICON system, with 119.61deg maximum in VICON and 54.37deg maximum in ARKit in the same joint. ARKit exposes smaller mean, SD, and ROM values. Similar observations apply to the inversion/eversion and rotation dimensions. The Pearson correlation-coefficients analysis shows that the ARKit and VICON measurements are strongly correlated, especially in the lower extremities. Even smaller changes in inversion/eversion are tracked in the squat motion's turning points, as shown in Figure 1b. Interestingly, the upper extremities' joint angle motion exhibit a considerably low correlation.

The ROM detected by ARKit is comparatively but reliably smaller than the ROM of VICON. This poses the question of whether an algorithm can be developed, which can approximate the ARKit values to the correct values provided by the VICON system. Enabling this would allow performing not only a quantitative motion analysis but also a qualitative analysis. A qualitative analysis would allow additional possibilities in motion analysis, mainly regarding feedback to prevent wrong motion.

Motion tracking using ARKit, especially regarding the lower extremity, seems to be a promising, light-weight approach. Even though the lab experiment shows that a quantitative assessment of exercise executions, e.g., for repetition counting, is feasible, the experiment was performed with a relatively limited number of participants and different exercises. A more extensive experiment is needed to gain further insights, which includes a more diverse set of participants and exercises. Through this, the limitations of the ARKit-based tracking need to be further investigated.

In case study 1, we showed that we can use ARKit data as reference values to recognize patterns in recordings. It remains open to which extent we can use reference data provided by ARKit to allow qualitative analysis of the motion, as proposed in case studies 2 and 3. The case studies served to explore the possibilities of ARKit-based motion tracking in health-related fields. Even though the first results show a high interest of potential app users and the lab experiment and data analysis expose such approaches' potential, the case studies need to be tested and validated in more extensive studies.

5. Conclusion

In this paper, we were able to show that even though ARKit exposes major inaccuracies in tracking the ROM, it is reliable in tracking the motion itself. Therefore, the ARKit framework can be used to assess physical exercise, recognize exercises, and count repetitions. ARKit enables various use cases for mobile applications, especially in the prevention and rehabilitation of diseases and injuries in humans. In this paper, we presented 3 different case studies. In case study 1, we classified exercises in goalkeeper training, which shows that we can identify patterns specific to exercises in the data provided by ARKit. In case study 2, we presented an algorithm that enables the creation and automatic tracking of dynamic body-weight exercises. In case study 3, we combined ARKit with an IMU sensor to enable a qualitative assessment of an exercise related to the golf swing.

ARKit seems to be a promising, light-weight alternative to well-established motion tracking systems. Its limitations and possibilities need to be further investigated.

References

[1] W. B. Kannel, P. Sorlie, Some Health Benefits of Physical Activity: The Framingham Study. *Arch Intern Med*, **139**(8) (1979), 857–861.

[2] A.P.A. Penedo and F. Ja, Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Curr Opin in Psychiatr*, **18** (2005), 189 – 193.

[3] D.E.R. Warburton, C.W. Nicol, S.S.D. Bredin, Health benefits of physical activity: the evidence. *CMAJ*, **174**(6) (2006), 801-809.

[4] J. Midtgaard, M. T. Baadsgaard, T. Møller, B. Rasmussen, M. Quist, C. Andersen, M. Rørth, and L. Adamsen. Self-reported physical activity behaviour; exercise motivation and information among Danish adult cancer patients undergoing chemotherapy. *Eur J of Oncol Nurs*, **13**(2) (2009), 116 – 121.

[5] L. E. Stephenson, D. G. Bebb, R. A. Reimer, S. N. Culos-Reed, Physical activity and diet behaviour in colorectal cancer patients receiving chemotherapy: associations with quality of life. *BMC Gastroenterol*, **9**(1) (2009).

[6] C. Visovsky. Muscle Strength, Body Composition, and Physical Activity in Women Receiving Chemotherapy for Breast Cancer. *Integr Cancer Ther*, **5**(3) (2006), 183–191.

[7] R.H. Sandmeier, Osteoarthritis and Exercise: Does Increased Activity Wear Out Joints?. *Perm J*, **4**(4) (2000), 26-28.

[8] M. Yahya, J.A. Shah, K.A. Kadir, Z.M. Yusof, S. Khan, A. Warsi, ""Motion capture sensing techniques used in human upper limb motion: a review"", *Sensor Rev*, **39**(4) (2019), 504-511.

[9] M. Field, Z. Pan, D. Stirling, F. Naghdy, Human motion capture sensors and analysis in robotics, *Ind Robot*, **38**(2) (2011), 163-171.

[10] E. van der Kruk, M.M. Reijne, Accuracy of human motion capture systems for sport applications; state-of-the-art review, *Eur J Sport Sci*, **18**(6) (2018), 806-819.

[11] O.D. Lara, M.A. Labrador. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun Surv Tut*, **15**(3) (2013), 1192-1209.

[12] M. N. Kamel Boulos, S.P. Yang. Mobile physical activity planning and tracking: a brief overview of current options and desiderata for future solutions. *mHealth*, **7**(13).

[13] S. Ishii, A. Yokokubo, M. Luimula, G. Lopez. ExerSense: Physical Exercise Recognition and Counting Algorithm from Wearables Robust to Positioning. *Sensors,* **21**(1), 91.

[14] S. Zelman, M. Dow, T. Tabashum, T. Xiao, M.V. Albert, Accelerometer-Based Automated Counting of Ten Exercises without Exercise-Specific Training or Tuning. *J Healthc Eng*, *2020*, 8869134.

[15] S. Basiratzadeh, E. D. Lemaire, N. Baddour, Augmented Reality Approach for Marker-based Posture Measurement on Smartphones, *Annu Int Conf IEEE Eng Med Biol Soc*, 2020, 4612-4615.

[16] J. Haladjian, D. Schlabbers, S. Taheri, M. Tharr, and B. Bruegge, Sensor-based Detection and Classification of Soccer Goalkeeper Training Exercises. *ACM Trans. Internet Things*, **1**(2) (2020).

[17] M. van Middelkoop, S.M. Rubinstein, T. Kuijpers, A.P. Verhagen, R. Ostelo, B.W. Koes, M.W. van Tulder, A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non‐specific low back pain. *Eur Spine J*, **20** (2011), 19–39.

[18] H. Ghasemzadeh, V. Loseu, R. Jafari. Wearable coach for sport training: A quantitative model to evaluate wrist-rotation in golf. *J Ambient Intell Smart Environ*, **1**(2) (2009), 173–184.

[19] U. Jensen., M. Schmidt, M. Hennig, et al., An IMU-based mobile system for golf putt analysis. *Sports Eng*, **18** (2015), 123–133.

[20] C.A. Ronao, S.B. Cho, Human activity recognition with smartphone sensors using deep learning neural networks, *Expert Syst Appl*, **59** (2016), 235-244.

5 Publication B: Evaluating 3D Human Motion Capture on Mobile Devices

This publication presents an extended evaluation of the mobile motion capture framework Apple *ARKit* as an example for mobile 3D motion capture frameworks against the motion capture system *Vicon*, serving as a gold standard.

Journal: Applied Sciences - Special Issue *Applied Biomechanics and Motion Analysis*
Number of Pages: 29
Type: Full Paper
Review: Peer Reviewed (2 Reviewers)

Summary

Problem. Optical motion capture systems enable assessing the quality of motion by tracking the positions and the angles of human body joints. Computer-vision-based frameworks further advanced these possibilities by providing markerless motion capture, which can even be performed on consumer-grade devices such as computers and mobile devices. These frameworks open up new fields of application for motion capture, for example, in the sector of sports and health, which is especially interesting on smart mobile devices. Most mobile motion capture frameworks rely on 2-dimensional tracking. However, 2D tracking is limited by the perspective of the observing camera. 3D motion capture frameworks promise to overcome these limitations and are enhanced by integrated smartphone sensors such as LiDAR. They consider all three movement planes independently of the camera angle but have only been evaluated to a limited extent so far. It remains unclear whether their tracking accuracy is sufficient for health and fitness applications. **Objective.** In this publication, we performed a laboratory experiment with 10 subjects to evaluate the accuracy of Apple *ARKit* against the *Vicon* system, the gold-standard system for optical motion capture. **Methods.** In the experiment, the subjects performed exercises using their body weight and were tracked by the *Vicon* system and two iPads running the *ARKit* software, filming from two positions. Both *ARKit*

data sets were compared against the *Vicon* results, considering the weighted Mean Absolute Error (wMAE) and the Spearman Rank Correlation Coefficient (SRCC). Factor analysis was performed using Welch ANOVA, Welch t-tests, and Logistic Regression to investigate influencing factors such as the camera position and the tracked exercise. **Results.** The results show a wMAE of $18.80° \pm 12.12°$ (ranging from $3.75° \pm 0.99°$ to $46.06° \pm 5.11°$). A mean SRCC of 0.76 was observed for the whole data set. The accuracy mainly depended on which angle was observed and which exercise was performed. **Conclusion.** While mobile 3D motion capture is a promising approach for sports and health applications, it still exhibits several limitations, such as the dependency of the tracking accuracy on which joint angle is observed and which exercise is performed. Additional limitations include the limited tracking of joint angles, which currently does not include the wrist and ankle angles, and the necessity to find a camera position where the joints of interest are well visible to the camera during the exercise. These limitations should be considered before application.

Contributions

The author of this dissertation is the lead/main author of this publication and contributed various aspects. She created the idea and concept of the research presented in this publication. She designed the experiment, implemented the application used for the *ARKit* recordings, and coordinated the data recordings during the experiment. She oversighted the data curation and analysis and was mainly responsible for the writing and creation of the visualizations of this publication. She performed the literature review. Moreover, she acquired funding and managed this research project as part of the MoMoTrack project, which was supported by a grant from the Software Campus Project through the German Federal Ministry of Education and Research.

applied sciences

Article

Evaluating 3D Human Motion Capture on Mobile Devices

Lara Marie Reimer [1,2,*], Maximilian Kapsecker [1,2], Takashi Fukushima [3] and Stephan M. Jonas [2]

1 Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany; max.kapsecker@tum.de
2 Institute for Digital Medicine, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany; stephan.jonas@ukbonn.de
3 Department of Sports and Health Sciences, Technical University of Munich, Georg-Brauchle-Ring 60/62, 80992 München, Germany; takashi.fukushima@tum.de
* Correspondence: reimer@tum.de

Featured Application: Mobile 3D motion capture frameworks can be integrated into a variety of mobile applications. Of particular interest are applications in the sports, health, and medical sector, where they enable use cases such as tracking of specific exercises in sports or rehabilitation, or initial health assessments before medical appointments.

Abstract: Computer-vision-based frameworks enable markerless human motion capture on consumer-grade devices in real-time. They open up new possibilities for application, such as in the health and medical sector. So far, research on mobile solutions has been focused on 2-dimensional motion capture frameworks. 2D motion analysis is limited by the viewing angle of the positioned camera. New frameworks enable 3-dimensional human motion capture and can be supported through additional smartphone sensors such as LiDAR. 3D motion capture promises to overcome the limitations of 2D frameworks by considering all three movement planes independent of the camera angle. In this study, we performed a laboratory experiment with ten subjects, comparing the joint angles in eight different body-weight exercises tracked by Apple ARKit, a mobile 3D motion capture framework, against a gold-standard system for motion capture: the Vicon system. The 3D motion capture framework exposed a weighted Mean Absolute Error of $18.80° \pm 12.12°$ (ranging from $3.75° \pm 0.99°$ to $47.06° \pm 5.11°$ per tracked joint angle and exercise) and a Mean Spearman Rank Correlation Coefficient of 0.76 for the whole data set. The data set shows a high variance of those two metrics between the observed angles and performed exercises. The observed accuracy is influenced by the visibility of the joints and the observed motion. While the 3D motion capture framework is a promising technology that could enable several use cases in the entertainment, health, and medical area, its limitations should be considered for each potential application area.

Keywords: human motion capture; mobile motion capture; optical motion capture; consumer electronics; mHealth; dHealth

check for updates

Citation: Reimer, L.M.; Kapsecker, M.; Fukushima, T.; Jonas, S.M. Evaluating 3D Human Motion Capture on Mobile Devices. *Appl. Sci.* **2022**, *12*, 4806. https://doi.org/10.3390/app12104806

Academic Editors: Rita M. Kiss and Alon Wolf

Received: 21 March 2022
Accepted: 6 May 2022
Published: 10 May 2022

1. Introduction

Human Motion Capture (HMC) is a highly researched field and covers the detection of all kinds of human motion, including movements of the whole body or smaller parts such as the face or hands [1]. In their publications from 2001 and 2006, Moesland et al. found more than 450 publications researching vision-based HMC and analysis [1,2], not considering HMC using different technologies such as inertial or magnetic sensors.

Traditional HMC systems are bound to an off-field setting [3,4] and are expensive in installation and operation [5,6], limiting their application to professional use cases. In their review of motion capture systems in 2018, van der Kruk and Reijne identified five types of motion capture systems: Optoelectronic Measurement Systems (OMS), Inertial Sensor Measurement Systems, Electromagnetic Measurement Systems (EMS), Ultrasonic

Localization Systems (ULS), and Image Processing Systems (IPS) [7]. They introduce OMS as the gold standard for motion capture [7]. Indeed, many studies [8–13] used OMS such as the Vicon motion capture system (Vicon, Oxford, UK) [14] or the Qualisys motion capture system (Qualisys AB, Göteborg, Sweden) [15] as reference measurement systems in their studies. OMS require multiple cameras or sensors around a subject and reflection markers on the subject's anatomical landmarks, which are then captured by the cameras or sensors. The Inertial Measurement Sensor Systems rely on Inertial Measurement Units (IMU), which are placed on the subject's body to capture motion and mapped onto a rigid-body model. Examples for IMU-based systems are the Xsens systems (Xsens Technologies B.V., Enschede, The Netherlands) [16] or Perception Neuron (Noitom Ltd., Miami, FL, USA) [17]. Through the traveling time of electromagnetic or ultrasonic waves between a tagged person and a base station, EMS and ULS track the position of the subject [7,18]. In contrast to the other systems, these systems allow tracking one or more subjects' positions, but do not capture joint kinematics [7]. While the described systems are well-validated systems for HMC, their complex setup and costs prevent them from application in mHealth applications. With the advancements in technology and machine learning, IPS became more relevant in human motion capture. IPS rely on video input and different machine learning approaches to detect specific body landmarks and capture human motion. Among the most researched systems is Kinect (Microsoft Corp., Redmond, WA, USA), which uses a combination of an RBG-camera and infrared sensors and can capture motion in 3-dimensional space [10,11]. However, the Kinect still requires a specialized setup for motion capture. The offer of IPS has been extended by recent advances in technology, such as enhanced sensors and processing units. These advances enable computer-vision-based motion capture on smartphones and tablets. These IPS systems offer new possibilities for HMC in mobile scenarios such as in mHealth applications. Examples for IPS software which can run on mobile devices are OpenPose (CMU, Pittsburgh, PA, USA) [19], ARKit (Apple Inc., Cupertino, CA, USA) [20], Vision (Apple Inc., Cupertino, CA, USA) [21], and TensorFlow Pose Estimate (Google, Mountain View, CA, USA) [22]. All of these IPS can be integrated into custom applications by developers. The detection of the human body and its position is realized through computer-vision algorithms, which can use Convolutional Neural Networks (CNNs) or Part Affinity Fields (PAFs) [23]. In most systems, a predefined humanoid model is then applied to estimate the shape and kinematic structure of the tracked person [2]. The algorithms deliver the joint coordinates in two or three dimensions for every video frame.

Moeslund et al. identified three main use cases for HMC: (1) surveillance of crowds and their behavior, (2) controlling software through specific movements or gestures or controlling virtual characters in the entertainment industry such as in movies, and (3) analysis of motion for diagnostics, for example in orthopedic patients or performance improvements in athletes [2]. While use case (1) focuses on tracking multiple subjects, (2) and (3) focus on capturing body motion of a single subject and thus require tracking of several parts of the human body. Especially use case (3) offers several applications of HMC, which are often limited to professional use cases such as gait analysis [24] or sports applications [7] due to the lack of a reliable, accessible, and low-priced solution in on-field settings.

In the sports and health sector, the usage of mobile applications has significantly increased in the past years [25,26]. Research has shown that such apps can positively impact their user's health and lifestyle [27]. However, most fitness and health apps only allow limited tracking and analysis of motion [28]. While smartphone-based motion capture promises a lightweight and consumer-friendly motion capture and analysis, the software systems have only been evaluated to a limited extent. Moreover, research has been focused on 2D systems. Several studies have shown that in 2D-motion analysis, the reliability and validity of the kinematic measurements are dependent on the performed task, which reliability is measured, video quality, and position of the recording device [8,13,29,30]. Especially the camera position influences the accuracy of tracked joint angles. A slightly different viewing angle already distorts the result of the joint angle, which is why triangulation with multiple devices is often performed to overcome the limitations of monocular camera setups [31]. Among mobile 2D motion capture systems, the OpenPose software is widely used and evaluated in several studies [19,23,30,32–37]. The results show that OpenPose delivers accurate biomechanical measurements, especially when tracking the joint trajectories. However, the compared joint angles differed significantly from the gold standard systems. D'Antonio et al. measured up to 9.9 degrees difference in the minima and maxima of the tracked joint angles during gait analysis [35], Nakano et al. experienced deviations of more than 40 mm in their study [37]. The measuerements can by improved by using multiple devices to calculate the body position in 3D as in the study by Zago et al. [30]. Mobile 2D motion capture systems have been recently complemented by 3D motion capture algorithms, which estimate the 3D joint positions based on 2D monocular video data [20,38–42]. They detect and calculate the body's joint coordinates in all three movement planes, making the motion capture more robust against the camera's viewing angle. Mobile 3D motion capture frameworks could overcome the limitations of 2D motion capture systems. Some of the 3D motion capture frameworks use additional smartphone sensors such as integrated accelerometers to determine the smartphone's position or depth sensors such as the integrated Light Detection and Ranging (LiDAR) depth sensor to additionally enhance the position detection of the human body [20,38,39]. The LiDAR data can be used to create a dense depth map from an RGB image through depth completion [43]. Among the most well-known mobile 3D motion capture systems is Apple ARKit, which released a body-tracking feature as part of their Software Development Kit (SDK) for developers in 2019 [20]. In contrast to other 3D motion capture frameworks, ARKit is free and easy to use, and widely accessible. On the latest devices, it uses the smartphone's IMUs and integrated LiDAR sensor to improve the measurements, promising enhanced mobile motion capture. However, only a few scientific studies have evaluated the accuracy of mobile 3D motion capture frameworks and ARKit in particular. Studies mostly focused on evaluating the lower extremity tracking of ARKit [44,45].

Due to the 3D calculations, ARKit is a promising IPS software that has the potential to enable new use cases for mobile HMC previously limited to traditional HMC systems. This research evaluated ARKit's performance against the Vicon system in a laboratory experiment in eight exercises targeting the whole body. We investigate the following two research questions:

- RQ 1: How accurate is ARKit's human motion capture compared to the Vicon system?
- RQ 2: Which factors influence ARKit's motion capture results?

2. Materials and Methods

2.1. Study Overview

To evaluate Apple ARKit's body tracking accuracy, we performed a laboratory experiment in which we compared the joint angles detected ARKit against the joint angles detected by the Vicon System for marker-based, optical motion tracking. In the experiment, ten subjects were instructed to perform eight different body-weight exercises with ten repetitions each, resulting in 80 recorded exercises.

During the exercises, the complete body of the subjects was recorded using the Vicon system and two iPads running ARKit from two different perspectives. All exercises were recorded simultaneously with the Vicon system and the two iPads. The study focused on comparing the motion capture data of each iPad against the data of Vicon to answer the underlying research questions. We calculated the weighted Mean Absolute Error (wMAE) and Spearman Rank Correlation Coefficient (SRCC) between the two systems in our data analysis. In addition, we performed factor analysis using ANOVA, *t*-tests, and logistic regression to quantify the impact of specific factors on the accuracy of the ARKit performance.

2.2. Participants

We included ten subjects ($n = 10$) in the study, six males and four females. Their age ranged from 22 to 31 years, with an average of 25.7 years. The subjects' height ranged between 156 cm and 198 cm with an average of 176 cm, and their weight was between 53 kg and 90 kg, with an average of 69.5 kg. All subjects had a normal body mass index between 20.4 and 25.5 (average: 22.7) and light skin color. All subjects were in good physical condition and did not have any orthopedic or neurological impairments.

2.3. Ethical Approval and Consent to Participate

The study was conducted according to the guidelines of the Declaration of Helsinki. The ethics proposal was submitted to and approved by the Ethics Committee of the Technical University of Munich on 19 August 2021—Proposal 515/21 S. All participants were informed about the process of the study upfront, and informed written consent was obtained from all subjects involved in the study. Due to the non-interventional character of this study, the risks involved for the study participants were low. We further minimized the risk through a sports scientist who supervised the physiologically correct execution of all exercises during the study, preventing the participants from performing potentially harmful movements.

2.4. Exercise Selection

Eight exercises were selected: Squat, Front Lunge, Side Squat, Single Leg Deadlift, Lateral Arm Raise, Reverse Fly, Jumping Jacks, and Leg Extension Crunch. The main objective of the exercise selection was to create a full-body workout to track all selected joints from different angles.

All exercises were tested for the suitability of tracking in both systems to ensure stable tracking of the angles. Both ARKit and the Vicon system exposed problems with the correct detection of exercises, where more extensive parts of the body were hidden from the cameras, for example, push-ups, and were therefore excluded. The testing was done in two steps: (1) We manually inspected the screen recording to see if the ARKit app model recognized the subject. (2) We checked the screen recording to whether the ARKit model overlayed with the subject's body parts during all parts of the exercise and whether the Vicon system could track all markers in the majority of recorded frames so that the full joint trajectory could be calculated.

Only if both requirements were fulfilled, we selected the exercise for the study. The final exercise selection included eight exercises. Their execution (E, see Figure 1) and targeting muscle groups (TMG) are explained in the following, and tracked joint angles (TJA) are explained in the following.

(I) Squat: (E:) The subject starts this exercise in an upright standing position. The subject squats down from the starting position by flexing the ankle, knee, and hip without movement compensations such as flexing the trunk and raising the heel. Each subject was asked to hold their arms stretched in front of the body. (TMG:) This exercise targets the lower body, especially the gluteus, quadriceps, hamstrings, and calves. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(II) Front Lunge: (E:) The starting position of the exercise is an upright standing with spreading legs front and back. The arms' position is the same as the squat. From the starting position, the subject goes down by flexing the ankle, knee, and hip in the front leg, flexing the knee and hip, and raising the heel in the back leg. (TMG:) This exercise targets lower body muscles, especially the gluteus, quadriceps, hamstrings, and calves. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

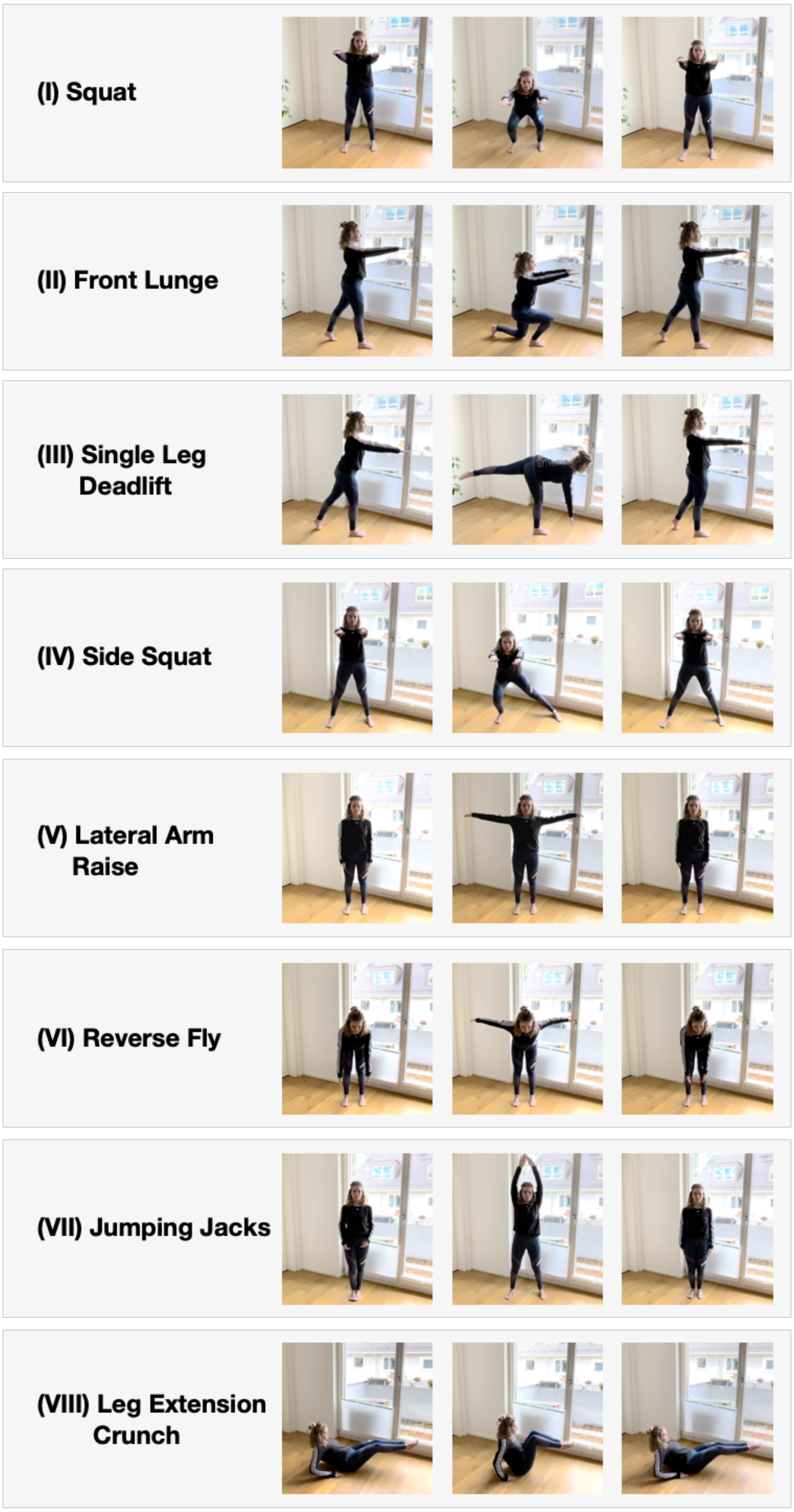


Figure 1. The execution of all eight exercises as seen from the frontally positioned iPad. The body orientation was chosen to maximize the visible parts of the body.

(III) Side Squat: (E:) The starting position of the exercise is an upright standing with spreading legs laterally. The arms' position is the same as the squat. From the starting point, the subject squats down with either side with either leg while the other leg is kept straight. (TMG:) This exercise targets similar muscle groups to squats, focusing on adductor muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(IV) Single Leg Deadlift: (E:) The starting position of the exercise is an upright standing with a single leg. The arms' initial position is the same as in the Squat. Th subject leans forward from the starting position by flexing the hip with minimum knee flexion. As the subject leans forward, the arms should be hung in the air. The other side of the leg in the air should be extended backward to maintain balance as the subject leans forward. (TMG:) The exercise targets lower body muscles, especially the hamstring and gluteal muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(V) Lateral Arm Raise: (E:) The subject starts the exercise in an upright standing position. Then, the subject laterally abducts the arms. (TMG:) The exercise targets upper body muscles, especially the deltoid muscles. (TJA:) The tracked joint angles include the left and right shoulder, and left and right elbow.

(VI) Reverse Fly: (E:) The subject leans forward with slight knee flexion and hangs the arms in the air in a starting position. The subject horizontally abducts the arms from the position without raising the upper body. (TMG:) The exercise targets upper body muscles such as the rhomboid, posterior deltoid, posterior rotator cuff, and trapezius muscles. (TJA:) The tracked joint angles include the left and right shoulder, and left and right elbow.

(VII) Jumping Jack: (E:) This exercise starts from an upright standing position. Then, the subject abducts both sides of the legs and arms simultaneously with a hop. (TMG:) This exercise targets lower body and upper body muscles, especially the gluteal and deltoid muscles. (TJA:) The tracked joint angles include the left and right shoulder, left and right elbow, left and right hip, and left and right knee.

(VIII) Leg Extension Crunch: (E:) The subject starts this exercise by sitting down on the ground with a backward lean of the upper body. The subject should place the hands on the ground to support the upper body as leaning backward. Then, the subject brings the legs in the air with knee and hip flexion. From the position, the subject extends the knee and hip horizontally on both sides together. (TMG:) This exercise targets core muscles, especially abdominal muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

2.5. Data Collection

We prepared the laboratory before the subjects arrived to ensure similar conditions for all recordings. Four tripods were positioned, each of them approximately three meters away from the area of the subjects' position to enable tracking of the entire body. Two tripods held an iPad Pro 11″ (2021 Model; Apple Inc., Cupertino, CA, USA), which were used to run the ARKit motion capture. Two other tripods were equipped with regular cameras to record videos of the experiment. One iPad and one camera were placed facing the subject's position frontally, the other iPad and camera were placed at an approximate angle of 30° facing the subject, as shown in Figure 2. The Vicon system (Nexus 2.8.2, Version 2.0; Vicon Motion Systems Ltd., Oxford, UK) was installed on the lab ceiling and configured to track the subjects' whole body.

We developed a protocol to guarantee a similar experiment execution for all participants. The experiment consisted of three phases: (1) the onboarding, (2) the explanation of the exercises, and (3) performing the exercises. During phase (1), the participants entered the lab. We explained the setup, and the participants signed the consent forms. In phase (2), a sports scientist explained each of the eight exercises and showed the participants how they are performed. The participants were asked to perform the exercises once under the supervision of the sports scientist to guarantee correct execution. The actual experiment was performed in phase (3). The participants performed ten repetitions of each exercise.

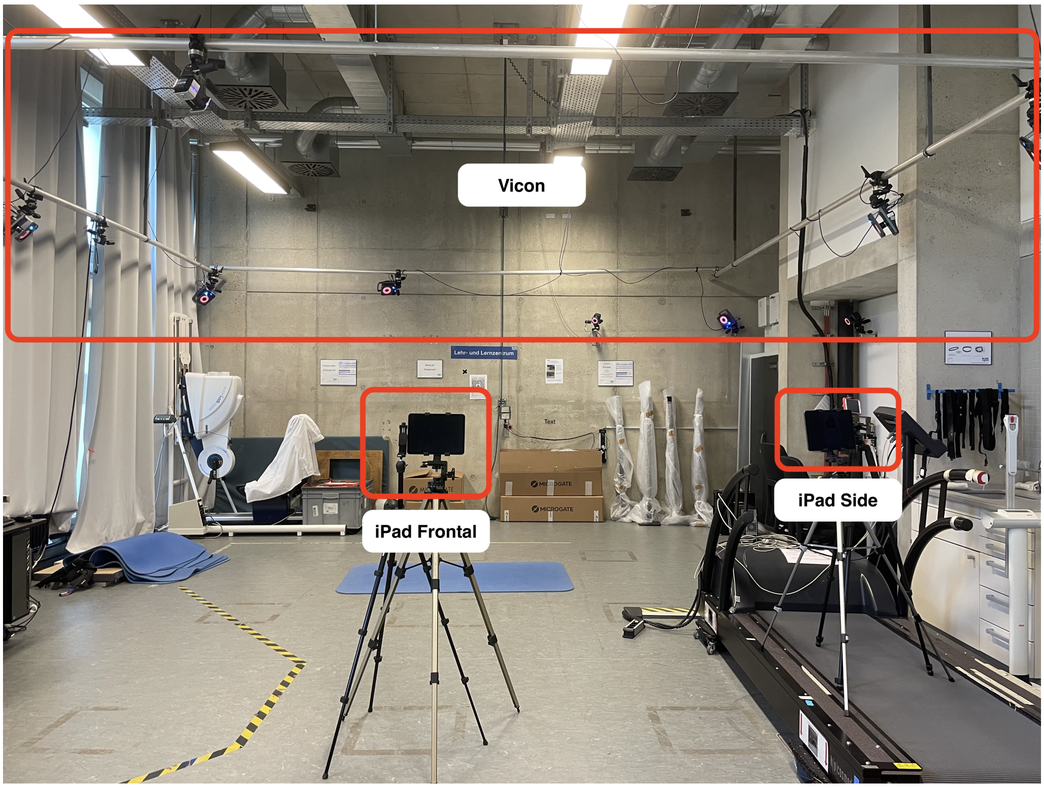


Figure 2. The experiment setup, showing the positioning of the recording devices and the subject.

2.5.1. Vicon Setup

The Vicon setup consisted of 14 infrared cameras. The setup included eight MX-T10-S cameras, four Vero v2.2 cameras, and two Bonita 10 cameras. All cameras were set to a sampling frequency of 250 Hz. We used the Nexus software (version 2.8.2) with the Full-Body Plug-in Gait marker placement model provided by Vicon Motion Systems, Ltd. [46] to capture the motion. A Vicon calibration wand was used to calibrate all the Vicon cameras and determine the coordinate system. Static calibration was done by capturing a subject performing a T-pose.

2.5.2. ARKit Setup

The ARKit setup included two iPad Pro 11″ 2021 with an M1 processor and an additional LiDAR sensor for depth information. Both iPads ran a custom-developed software based on the ARKit 5 framework provided by Apple Inc., which was used for extracting the motion capture information from the iPads' sensors. Both iPads recorded the motion capture data independently and were not synchronized. The motion capture data included the timestamp of the detection, the performed exercise, and the three-dimensional, positional information of 14 body joints. These data were later used to calculate the joint angles. All joint coordinates are given relatively to the pelvis center, which serves as the origin of ARKit's coordinate system. ARKit differentiates between bigger joints, which are actively tracked, and calculated joints, which are smaller joints such as the toes and fingers. We decided only to include actively tracked joints in our comparison, as previous tests showed that the calculated position of the smaller joints and their related angles rarely change. The ARKit data were recorded with a default sampling frequency of 60 Hz. However, the sampling frequency of ARKit is variable, as ARKit internally only updates the joint positions when a change is detected. This means that if a subject is standing still, fewer data points are received from ARKit and more when the subject is moving fast. As the toe and finger joints are calculated by ARKit and not actively recognized, we limited the comparison to the actively tracked joints: shoulders, elbows, hips, and knees.

2.5.3. Data Export

After each recorded subject, the collected motion data were exported from the three systems: the frontally positioned iPad (iPad Frontal), the iPad set in a 30° Side Angle (iPad Side) (Figure 2), and the Vicon system. The motion data were stored in CSV files and included the joint center coordinates for each detected frame for the three systems separately. The ARKit data were exported in one file per iPad, resulting in two CSV files per subject. For the Vicon system, each exercise was stored in a separate CSV file. In addition, an XCP file was exported from the Vicon system, which contained meta-information about the cameras, including the start and end timestamps of each recording.

Due to export problems, the upper body joint coordinates of the iPad Side were only included for three of the ten subjects. The Vicon system could not track each joint coordinate throughout the whole exercise due to hidden markers, leading to gaps in the exported data. Smaller gaps were compensated during the Data Analysis, whereas more significant gaps led to the exclusion of the respective angle.

2.6. Preprocessing & Data Analysis

The basis for the data analysis part is 220 files, 22 for each subject. It contains two comma-separated value (CSV) files from the respective ARKit systems (frontal and side view) and ten CSV files from the Vicon system, which records each exercise in a separate file. The remaining ten files are given in the XCP format, which contains the relevant metadata of the Vicon system, such as camera position, the start time, and the end time of the data acquisition process. The following preprocessing steps are performed for each subject to merge all files into a data frame for further analysis.

The Vicon and ARKit data are modified to fit a matrix-like structure in which the rows represent time and columns the joints. Augmentation enhances the data with information such as the timestamp, subject, exercise, and in the case of ARKit, whether the values were recorded frontal or lateral.

The Sections 2.5.1 and 2.5.2 explain different sampling rates for the systems and the non-equidistant sampling rate of ARKit (57 Hz on average). It motivates to evaluate strategies to merge the system's data based on the timestamp. Vicon samples the data at a frequency of 250 Hz and implies a maximum of 2 ms distance for a randomly chosen timestamp. Due to this maximal possible deviation, the nearest timestamp is the criterion for merging the Vicon data onto the ARKit data.

The Vicon system records absolute coordinates, while the ARKit system provides normalized coordinates relative to the center of the hip. It still allows for comparing angles since they are invariant under scaling, rotating, translating, and reflecting the coordinate system. Accordingly, the adjacent three-dimensional joint coordinates extraction calculates the angles of interest (AOI). An angle θ is determined by three joints $A, B, C \in \mathbb{R}^3$ or associated vectors $\vec{v_1} = A - B$ and $\vec{v_2} = C - B$ given the formula

$$\theta = \arccos \frac{v_1 \cdot v_2}{||v_1||_2 ||v_2||_2}$$

The data reveal a time lag which leads to a misalignment between the Vicon and ARKit angles along the time axis. Accordingly, the related time series require shifting with the objective to maximize the mutual Pearson correlation coefficient. The shift operation is subjected to a maximum of 60 frames to each side. It includes the assumption that the time series of the two systems match best if they exhibit similar behavior in their linear trends. Figure 3 shows two examples of misaligned time series on the left and the result of the shift on the right. The time series alignment is performed brute force and individually for any combination of view, subject, exercise, and AOI. The procedure outputs 1048 ARKit-Vicon time series pairs, 634 for the comparison Vicon—iPad Frontal, and 414 for the comparison Vicon—iPad Side. The number does not correspond to $2 \times 10 \times 8 \times 8$ = 1280 pairs due to the missing ARKit recordings of the upper body joints for lateral recording.

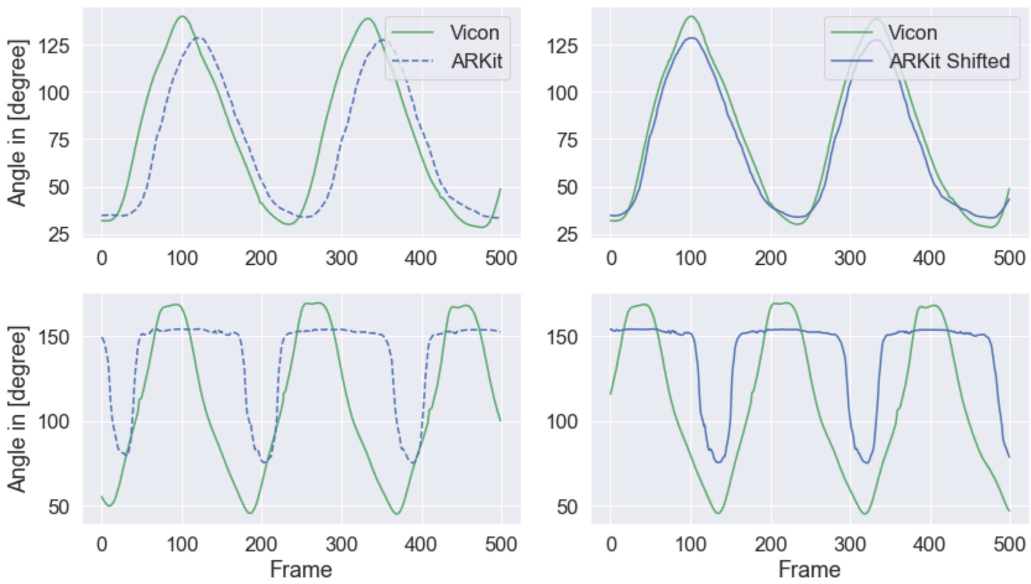


Figure 3. Shift of the data.

Computing two metrics validates the angle similarity of the systems for each pair of time series, the mean absolute error (MAE) and the non-parametric Spearman's rank correlation coefficient (SRCC). The obtained MAE and SRCC values of the 1048 time series are aggregated according to predefined grouping criteria, such as exercise, angle, or view. Calculating the sample size's weighted mean and standard deviation (std) defines a grouping operation for the MAE (Table 1). SRCC values require first a transformation to a normally distributed random variable using the Fisher z-transformation

$$z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) \quad (1)$$

where r is the SRCC. It constitutes the prerequisite to applying the averaging operation along with the variables. The result is again a normally distributed variable that needs back transformation into the correlation space using the inverse of (1).

Table 1. The aggregated wMAE values for all joint angles.

Angle	wMAE
leftElbow	$24.0° \pm 17.43°$
leftHip	$16.91° \pm 10.67°$
leftKnee	$16.61° \pm 7.47°$
leftShoulder	$20.01° \pm 14.89°$
rightElbow	$20.0° \pm 15.32°$
rightHip	$20.17° \pm 11.25°$
rightKnee	$17.57° \pm 7.25°$
rightShoulder	$17.39° \pm 12.18°$

A drawback of the MAE is the lack of interpretation regarding systematic over- or underestimation of the angles. The mean error (ME), which is the average of the time series pair's difference, can conclude the occurrence of bias but at a granular level, for example segments of the exercise. However, aggregation of the ME is prone to involve effects such as error cancellation. The ratio of ME and MAE, for instance $\frac{\text{ME}}{\text{MAE}}$, draws insights into the occurrence of systematic bias (Figure A4). A value close to ± 1 implies less tendency of ARKit to fluctuate around the Vicon's angle estimation, for example either under-, perfect- or overestimation takes place. Values nearby zero indicate the ME's cancellation effect (over- and underestimation) but require further analysis, such as the difference between MAE and ME, for conclusions.

One-way analysis of variance (ANOVA) is performed to quantify the effects of the categorical metadata such as angle (fixed effect), exercise (fixed effect), and subject (random effect), on the continuous variable MAE. The random effect was taken into account performing one-way ANOVA using a random effects model. The distribution of MAE shows a divergence towards the normal distribution, which is one of the requirements in ANOVA. However, research verified robustness in violating this assumption in certain bounds [47]. A logarithm (basis 10) transform on the MAE variable ensures stronger normalization (Appendix A, Figure A1). In particular, it makes the model multiplicative and more robust to dispersion. The visual inspection of histograms reveals a lack of homogeneous intergroup variance and motivates to apply Welch's ANOVA. Finally, the Games-Howell post-hoc test [48] compares the individual categorical factors for significant results (here defined as an effect size larger than 0.1).

Besides view (frontal or side), the binary independent variables are the body segment of the angle (lower or upper) and information on the movement of the pelvis. The latter is declared as the variable center moved and indicates whether the proper execution of the exercise involves the movement of the pelvic's center, the origin of the ARKit coordinate system. To quantify the binary variables' effect, we fitted a logistic regression model based on the MAE and applied Welch's *t*-test. The results, including β coefficient, R^2 *p*-value, and confidence interval, are compiled in a table.

Assumptions about the data are made and can restrict the interpretation of the results. A more detailed outline of this topic is given in the limitations section (Section 5.8).

3. Results

3.1. Weighted Mean Absolute Error

3.1.1. Aggregated Results

The data analysis exposed a wMAE of $18.80^\circ \pm 12.12^\circ$ degrees for all angles in the whole data set. The wMAE across all exercises, views, and angles is visualized in Figure 4 to enable more profound insights into the performance based on exercises and joint angles.

The data exposed high differences in the detected error rates with the wMAE ranging between $3.75^\circ \pm 0.99^\circ$ (Lateral Arm Raise, Left Elbow, Side) and $47.06^\circ \pm 5.11^\circ$ (Side Squat, Left Elbow, Side), depending on the performed exercise and observed joint. To generate better insights into the different factors, we aggregated the wMAE by angle, performed exercise, view, and subject.

Considering the aggregated wMAE for the individual joints (Table 1), the mean value ranged between $16.61^\circ \pm 7.47^\circ$ for the left knee up to $24.00^\circ \pm 17.43^\circ$ for the left elbow. The left hip exposed a wMAE of $16.91^\circ \pm 10.67^\circ$, followed by the right shoulder with a wMAE of $17.39^\circ \pm 12.18^\circ$ and the right knee with a value of $17.57^\circ \pm 7.25^\circ$. The right elbow had a wMAE of $20.00^\circ \pm 15.32^\circ$, the left shoulder $20.01^\circ \pm 14.89^\circ$ and the right hip $20.17^\circ \pm 11.25^\circ$.

The observed wMAE differed between the exercises, with the Lateral Arm Raise ($9.56^\circ \pm 6.13^\circ$), Jumping Jacks ($10.09^\circ \pm 3.81^\circ$), Single Leg Deadlift ($11.35^\circ \pm 5.04^\circ$), Reverse Fly ($15.80^\circ \pm 8.5^\circ$), Leg Extension Crunch ($18.15^\circ \pm 8.21^\circ$), and Front Lunge ($18.19^\circ \pm 8.98^\circ$) exposing significantly lower error rates than the Side Squat ($30.49^\circ \pm 12.73^\circ$) and the Squat ($33.79^\circ \pm 10.25^\circ$) (Table 2).

When only considering the targeted joints, the wMAE ranged between $3.75^\circ \pm 0.99^\circ$ (Lateral Arm Raise, Left Elbow, Side View) and $38.41^\circ \pm 6.66^\circ$ (Squat, Right Hip, Frontal View). The exercises Lateral Arm Raise, Reverse Fly, and Single Leg Deadlift performed best with wMAE values below 15.00° in the relevant joints for the respective exercises. The wMAE of Jumping Jacks, Front Lunge, and Leg Extension Crunch remained below 25.00° across the targeted joints. The Squat and Side Squat Exercises exposed error rates of up to 38.41° in the targeted joints and thus performed worst in the experiment.

Exercise	Angle View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	35.26 ± 8.06	23.75 ± 7.44	8.39 ± 4.46	12.36 ± 3.23	10.83 ± 4.12	20.14 ± 6.82	13.99 ± 2.76	13.97 ± 3.90
	Side	36.66 ± 2.49	33.68 ± 4.20	14.7 ± 0.62	17.6 ± 2.51	12.54 ± 3.12	22.26 ± 3.19	17.57 ± 4.07	17.96 ± 2.93
Jumping Jacks	Frontal	7.66 ± 3.01	7.02 ± 2.50	6.60 ± 1.27	7.58 ± 1.50	8.4 ± 1.57	8.33 ± 1.86	15.15 ± 2.56	15.02 ± 2.00
	Side	9.14 ± 1.07	9.32 ± 0.30	6.49 ± 1.18	7.77 ± 0.42	9.39 ± 1.97	9.93 ± 2.58	14.9 ± 2.12	13.94 ± 1.80
Lateral Arm Raise	Frontal	7.51 ± 3.35	7.36 ± 2.67	6.65 ± 1.34	6.81 ± 1.97	5.01 ± 2.67	4.80 ± 2.7	17.67 ± 4.17	17.61 ± 3.32
	Side	3.75 ± 0.99	5.43 ± 1.36	5.5 ± 0.39	5.69 ± 0.42	4.93 ± 2.14	6.84 ± 3.28	17.76 ± 4.39	17.44 ± 3.86
Leg Extension Crunch	Frontal	18.94 ± 6.77	19.49 ± 5.99	31.98 ± 10.20	17.39 ± 5.80	10.86 ± 2.96	11.49 ± 3.86	14.81 ± 4.18	16.67 ± 3.07
	Side	19.78 ± 6.42	26.07 ± 10.73	36.33 ± 4.94	21.29 ± 4.66	14.0 ± 4.81	16.44 ± 6.24	17.18 ± 3.88	20.17 ± 6.00
Reverse Fly	Frontal	8.90 ± 3.91	10.27 ± 3.98	11.84 ± 4.02	10.63 ± 3.80	25.08 ± 6.27	27.73 ± 6.31	15.54 ± 7.54	15.01 ± 7.23
	Side	7.89 ± 2.10	14.85 ± 5.22	14.72 ± 3.68	8.69 ± 2.99	22.32 ± 7.14	20.98 ± 7.48	15.37 ± 8.10	11.98 ± 7.05
Side Squat	Frontal	46.73 ± 14.20	42.48 ± 14.29	41.82 ± 9.72	36.21 ± 8.84	22.62 ± 7.63	36.41 ± 5.21	16.50 ± 6.13	26.48 ± 4.13
	Side	47.06 ± 5.11	27.74 ± 4.13	30.04 ± 2.17	23.67 ± 1.60	26.19 ± 8.28	30.25 ± 3.95	16.17 ± 5.88	21.54 ± 2.76
Single Leg Deadlift	Frontal	21.07 ± 4.82	8.59 ± 4.60	11.78 ± 2.77	10.82 ± 2.96	10.57 ± 3.33	14.37 ± 4.20	8.66 ± 2.47	9.25 ± 3.49
	Side	13.26 ± 1.06	6.11 ± 0.98	9.39 ± 1.36	7.13 ± 1.59	12.28 ± 4.96	14.91 ± 4.61	8.17 ± 2.85	10.0 ± 3.40
Squat	Frontal	44.35 ± 17.93	37.85 ± 16.51	39.23 ± 9.04	37.49 ± 8.70	35.37 ± 6.46	37.41 ± 6.66	29.94 ± 4.42	30.36 ± 3.72
	Side	45.40 ± 10.51	31.86 ± 9.57	36.56 ± 5.05	27.77 ± 3.06	32.04 ± 4.63	30.05 ± 4.84	27.98 ± 3.31	23.73 ± 2.00

Figure 4. Pivot Table of the weighted Mean Absolute Error (wMAE) in degrees distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker green color referring to a lower error rate and darker orange color referring to higher error rates.

When only considering the targeted joints per exercise, the wMAE was reduced for all exercises except the Jumping Jacks, where the wMAE remained the same (Table 2).

Table 2. The wMAE values for all exercises when considering all angles and only the targeted angles per exercise.

	All Angles	Targeted Angles
Front Lunge	$18.19° \pm 8.98°$	$16.17° \pm 5.48°$
Jumping Jacks	$10.09° \pm 3.81°$	$10.09° \pm 3.81°$
Lateral Arm Raise	$9.56° \pm 6.13°$	$6.66° \pm 2.41°$
Leg Extension Crunch	$18.15° \pm 8.21°$	$15.14° \pm 5.34°$
Reverse Fly	$15.80° \pm 8.5°$	$10.67° \pm 4.31°$
Side Squat	$30.49° \pm 12.73°$	$24.56° \pm 8.63°$
Single Leg Deadlift	$11.35° \pm 5.04°$	$10.91° \pm 4.41°$
Squat	$33.79° \pm 10.25°$	$30.93° \pm 6.19°$

The difference between the view of the recording device was smaller than the observed differences between the exercises, with an wMAE of $17.91° \pm 9.68°$ for the side view and $19.35° \pm 13.38°$ for the frontal view.

When considering the different subjects, the observed wMAE was relatively consistent among the individuals, with mean values ranging from $16.20° \pm 9.44°$ to $22.32° \pm 17.08°$.

3.1.2. Bias of the ARKit System

For detecting a possible bias of over- and underestimation of the ARKit data, we investigated the ME and the ratio of ME/MAE. The aggregated results of the ME/MAE ratio exhibits only seven values below 0.1 for the exercise—angle—view configurations (Appendix B Figure A3 for the ME, Appendix B Figure A4 for ratio ME/MAE). In 4 of these cases, the wMAE is above 10°: Front Lunge—left hip—Frontal, Jumping Jacks—left knee—Frontal, Jumping Jacks—right knee—Frontal, and Leg Extension Crunch—left elbow—Frontal. Most other values remain relatively close to 1 or −1.

3.2. Spearman Rank Correlation

The whole dataset exposed a mean Spearman Rank Correlation Coefficient of 0.76. The *p*-value was below 0.01 for 1019 of the 1048 exercises. A detailed overview of the individual SRCCs, including the standard deviation for the exercises, is visualized in Figure 5.

Exercise	Angle / View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	0.22	0.16	0.67	0.65	0.49	0.93	0.91	0.95
	Side	-0.13	-0.20	0.59	0.62	0.63	0.97	0.92	0.97
Jumping Jacks	Frontal	0.36	0.25	0.91	0.90	0.43	0.42	0.32	0.40
	Side	0.63	0.47	0.93	0.91	0.32	0.66	0.43	0.70
Lateral Arm Raise	Frontal	0.79	0.82	0.96	0.96	0.54	0.25	0.22	0.13
	Side	0.78	0.68	0.99	0.96	0.61	0.45	0.26	0.17
Leg Extension Crunch	Frontal	0.55	0.69	0.44	0.85	0.94	0.92	0.92	0.90
	Side	0.26	0.68	0.32	0.81	0.90	0.85	0.93	0.89
Reverse Fly	Frontal	0.45	0.47	0.87	0.84	0.80	0.79	0.50	0.53
	Side	0.45	0.33	0.80	0.84	0.74	0.76	0.43	0.49
Side Squat	Frontal	-0.05	-0.08	0.56	0.65	0.90	0.90	0.63	0.93
	Side	0.13	-0.03	0.25	0.52	0.91	0.98	0.63	0.97
Single Leg Deadlift	Frontal	0.33	0.69	0.86	0.89	0.95	0.77	0.74	0.51
	Side	0.55	0.63	0.97	0.97	0.94	0.70	0.78	0.49
Squat	Frontal	-0.06	-0.15	0.71	0.70	0.75	0.79	0.84	0.89
	Side	-0.27	0.05	0.77	0.60	0.88	0.95	0.90	0.97

Figure 5. Pivot Table of the average Spearman Rank Correlation Coefficients (SRCC) distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker green color referring to a higher positive correlation and darker orange color referring to a higher negative correlation.

The SRCC varied between the tracked angles with a range of -0.27 to 0.99 as mean values per exercise and angle as displayed in Figure 5. When considering the results aggregated per joint angles (Table 3), all negative correlations were observed for the elbow angles (left elbow 0.36, right elbow 0.42) in both iPad views, with the side view performing worse than the frontal view. The shoulder angles exposed a mean SRCC of 0.81 for both shoulders. Knee and hip joints were also tracked with moderate SRCC values (left hip: 0.82, right hip: 0.84, left Knee: 0.75, right knee: 0.81).

Table 3. The aggregated SRCC values for all joint angles.

Angle	SRCC
leftElbow	0.36
leftHip	0.82
leftKnee	0.75
leftShoulder	0.81
rightElbow	0.42
rightHip	0.84
rightKnee	0.81
rightShoulder	0.81

While the SRCCs differed between the exercises, all of them exposed moderate linear correlations with values above 0.5 (Table 4). The Leg Extension Crunch showed a correlation of 0.84. Front Lunge correlated with 0.80, followed by the Single Leg Deadlift with an SRCC of 0.79. The Squat and Side Squat exercises showed a correlation of 0.78. The SRCC of the

Lateral Arm Raise was 0.68, and the SRCC of the Reverse Fly was 0.67. The Jumping Jacks performed worst with a correlation of 0.60.

Similar to the wMAE, considering only the relevant joints for the specific exercises positively influenced the SRCCs of all exercises except for the Jumping Jacks, where it remained the same, and the Single Leg Deadlift, where it was reduced by 0.01 (Table 4).

Table 4. The average SRCC values for all exercises when considering all angles and only the targeted angles per exercise.

	SRCC All Angles	SRCC Targeted Angles Only
Front Lunge	0.80	0.91
Jumping Jacks	0.60	0.60
Lateral Arm Raise	0.68	0.91
Leg Extension Crunch	0.84	0.91
Reverse Fly	0.67	0.69
Side Squat	0.78	0.91
Single Leg Deadlift	0.79	0.78
Squat	0.78	0.89

Comparing the two positions of the iPads, the side view performed slightly better than the frontal view, with SRCCs of 0.80 and 0.73, respectively.

Similar to the wMAE, the SRCC is relatively consistent across the recorded subject, with values between 0.72 and 0.82.

3.3. Factor Analysis

3.3.1. ANOVA Analysis

To further investigate the influence of the observed exercise, angle, and subject on the performance of ARKit, we performed a Welch ANOVA factor analysis on the Mean Absolute Error for the factors *Exercise* and *Angle* and a random effects model for the factor *Subject*. The MAE exhibited a high dependency on the observed exercise with an effect size of $\eta^2 = 0.51$ ($p = 0.00$). It did not expose a dependency on the observed angle ($\eta^2 = 0.03$, $p = 0.00$). The random effects model analysis did not exhibit an influence of the subject, with 0.29% of the variance explained by the subject (Table 5).

Table 5. The results of the Random Effects ANOVA.

Random Effects			
Groups	**Name**	**Variance**	**Std. Dev.**
Subject	(Intercept)	0.001312	0.03622
Residual		0.458310	0.67699
Fixed Effects			
	Estimate	Std. Error	*t* value
(Intercept)	2.70803	0.02389	113.4

To further investigate the influencing factors of the performed exercise in the MAE, we performed a Post-hoc analysis using the Games-Howell test (Appendix C, Table A1). The exercise analysis exhibits significant differences between 20 of the 28 exercise pairs.

3.3.2. Welch *t*-Test Analysis

All binary influencing factors of the MAE were analyzed using Welch's *t*-test (Table 6). The results of the *t*-test showed a dependency on the pelvic center movement ($cohen - d = 0.82$, $power = 1.00$, $p = 0.00$). No dependency was measured for the view ($cohen - d = 0.01$, $power = 0.06$, $p = 0.82$), and whether the measured angle is a lower body angle ($cohen - d = 0.01$, $power = 0.05$, $p = 0.88$).

Table 6. The results of the Welch *t*-test Analysis.

T	dof	Alternative	*p*-Value	CI95%	Cohen-d	BF10	Power	Response	Categorical
−0.22	966.81	two-sided	0.82	[−0.09, 0.07]	0.01	0.073	0.06	LogMAE	View
−0.15	725.74	two-sided	0.88	[−0.1, 0.08]	0.01	0.072	0.05	LogMAE	LowerBody
−13.20	1045.97	two-sided	0.00	[−0.59, −0.44]	0.82	3.266×10^{33}	1.00	LogMAE	CenterMoved

3.3.3. Logistic Regression Analysis

In addition to the *t*-test, we applied logistic regression to the three variables *View*, *LowerBody*, and *CenterMoved* (Table 7). The logistic regression model for the *LowerBody* shows a slight effect with $\beta - coef = 0.0684$ ($p = 0.00$). The model exposed a $Pseudo - R^2$ of 0.165. While the *View* model exposed no significant effect ($\beta - coef = 0.0141$, $p = 0.00$), the fitness of the model is low ($Pseudo - R^2 = -0.019$). The *CenterMoved* variable showed no effect ($\beta - coef = 0.0018$, $p = 0.575$). Similar to the *View* variable, the $Pseudo - R^2$ of 0.000 indicated bad fitness of the model to explain the data.

Table 7. The results of the logistic regression.

Variable	β-$coef$	std	z	P > |z|	[0.025	0.975]	$Pseudo$-R^2
View	0.0141	0.003	4.329	0.000	0.008	0.020	−0.019
Lower Body	0.0684	0.005	13.374	0.000	0.058	0.078	0.165
Center Moved	0.0018	0.003	0.561	0.575	−0.004	0.008	0.000

4. Findings

While the results showed that ARKit is generally capable of tracking human body motion, the accuracy of the joint angles is highly variable and dependent on several factors, especially the performed exercise.

4.1. RQ 1: How Accurate Is ARKit's Human Motion Capture Compared to the Vicon System?

To answer RQ 1, we investigated both the wMAE and the SRCC of the experiment data. A wMAE of 0° and an SRCC of 1.0 would represent a perfect accuracy of ARKit's human motion capture. The ARKit data showed a MAE of 18.80° and an average SRCC of 0.76 for the whole data set, with variations when examining different joints and exercises. Based on the results of the ANOVA analysis, the accuracy mainly depends on the observed angle and exercise. However, the accuracy could be influenced by other additional factors which were not specifically targeted by the performed experiment. Remarkably, ARKit was able to achieve an almost perfect correlation and accuracy for some exercise executions in specific angles (Figure 6). In many cases, the movement pattern is recognizable in the ARKit data. Still, the amplitude is reduced, or a baseline drift on the y-axis is observable (Figure 7, which explains the good correlation but relatively high wMAE values. In some cases, the ARKit data exhibits high wMAE values and no or even a negative SRCC. These effects often occurred in the elbow joints, especially when the lower body joints moved and the upper body joints were held straight, such as in the Squat or Side Squat exercises. In this situation, ARKit often failed at detecting the movement correctly (Figure 8), which is visible both in the high wMAE and the low to negative correlation values for the elbow angles. In general, the accuracy was lower in those exercises where the root position did not remain stable, including the Front Lunge, Side Squat, and Squat exercises. The results of the factor analysis further confirmed these results.

To investigate whether a systematic baseline drift can be observed in the ARKit data, we aligned the ARKit and Vicon data via cross-correlation. We measured the y-axis offset (Figure 9). As the offset was normally distributed around 0, no systematic baseline drift was present in the recorded data set, indicating that other factors cause shifts.

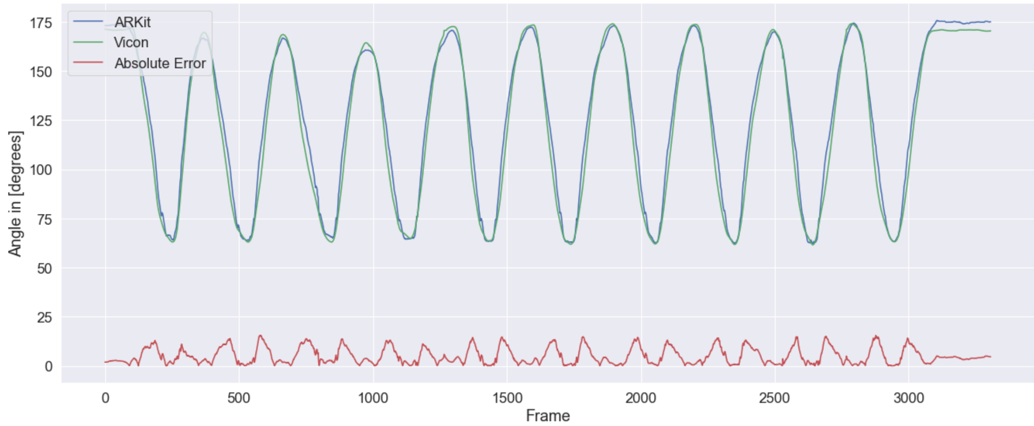


Figure 6. Left hip angle of one of the subjects in the Single Leg Deadlift exercise in degrees, which shows a nearly perfectly overlapping curves of the ARKit and Vicon data.

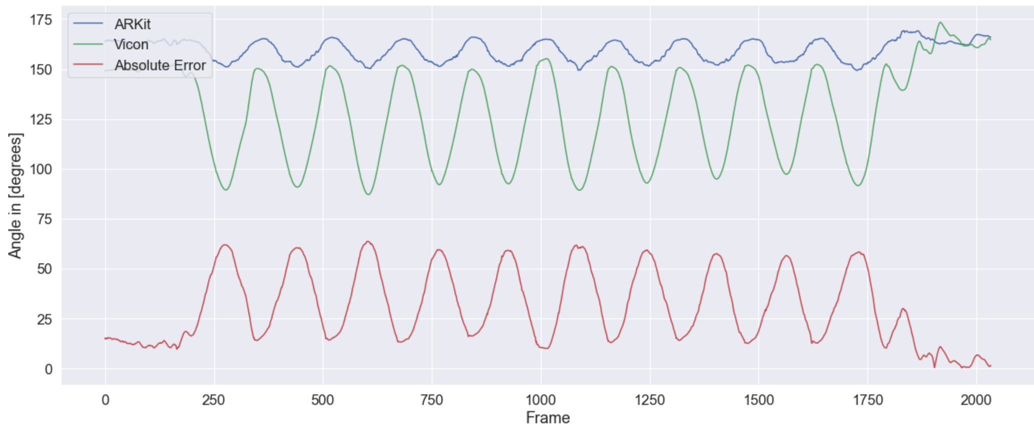


Figure 7. Left hip angle of one of the subjects in the Side Squat exercise in degrees. The plot shows that while the motion pattern is visible in both recordings, ARKit exposes a reduced amplitude and a shift on the y-axis.

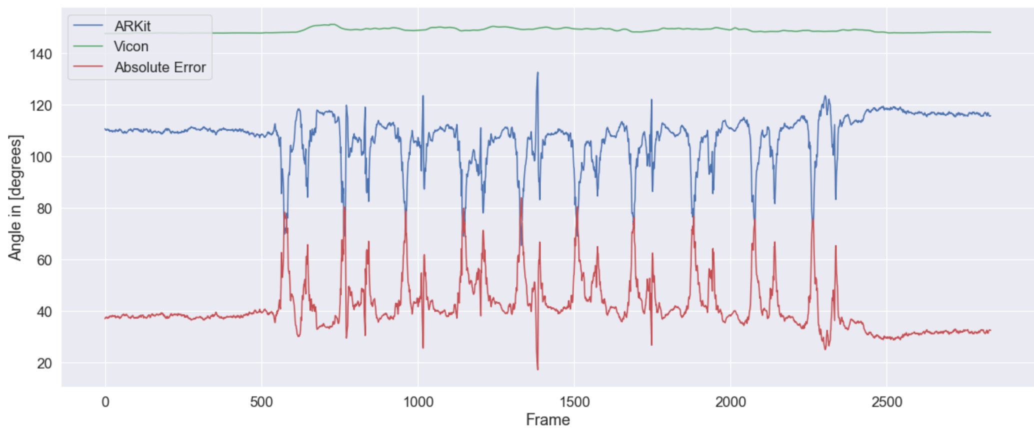


Figure 8. Right elbow angle of one of the subjects in the Squat exercise in degrees, which shows bad tracking quality with a lot of noise compared to the Vicon data.

Finding 1: ARKit is able to track the general progression of a movement with good accuracy but with significant deviations from the actual values measured by the Vicon system. The performance is influenced by external factors such as the performed motion.

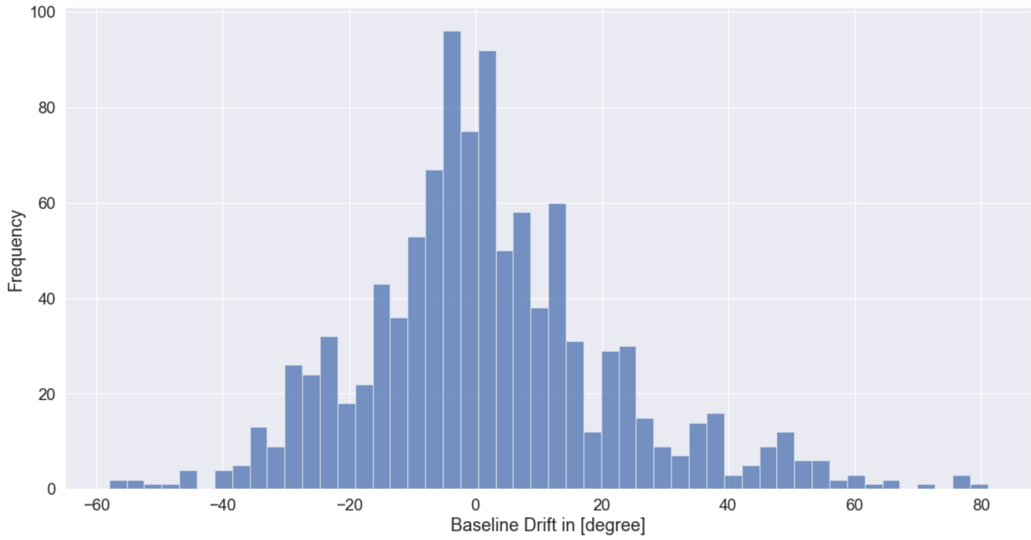


Figure 9. Results of the baseline drift analysis of the ARKit data. This is computed by minimizing the MAE by shifting the ARKit data vertically. The results show a normal distribution around 0, thus indicating no systematic baseline drift of the ARKit results.

4.2. RQ 2: Which Factors Influence ARKit's Motion Capture Results?

We performed factor analysis using Welch ANOVA, *t*-test analysis, and logistic regression on the dependent variable *MAE* to answer RQ2.

The MAE depended on the performed exercise. This dependency is visible when inspecting the respective boxplots of the MAE (Figure 10). Especially both Squat exercises (Squat, Side Squat) show significantly higher mean values than the other exercises. This observation is supported by the post-hoc analysis results of the ANOVA results. The logistic regression indicated an additional small influence of whether upper or lower body angles are considered. While the *t*-test showed an additional effect on whether the pelvic's center was moved during an exercise, this effect was not visible in the logistic regression. The impact of this factor remains inconclusive.

Finding 2: The factor analysis results show that the accuracy of ARKit's human motion capture mainly depends on the performed exercise.

While there is a slight difference between the frontal and side view data for both the wMAE and the SRCC, this difference is comparably small. The results of the side view show a 1.44° difference of the wMAE and a difference in the SRCC of 0.07, with the side view performing slightly better than the frontal view. These findings are supported by the factor analysis results, where no dependency of the view was measured. It also needs to be considered that the upper body angles in the side view only contained data of three subjects due to export problems, limiting the comparison's explanatory power.

Another aspect of the device's position influence is the visibility of specific body parts. Limited visibility of body joints, such as the left side of the body in the Front Lunge, Single Leg Deadlift, and Leg Extension Crunch, or the elbow joints in the Side Squat and Squat, is associated with a higher wMAE and worse correlation results, especially in the left elbow joint. Hidden joints often led to ARKit confusing the left and right body side for the respective joints, which caused unexpected peaks in the recorded data (Figure 11). The tracking of the upper body joints worked significantly better when other body parts did not hide them, as in the remaining three exercises Jumping Jacks, Lateral Arm Raise, and Reverse Fly.

Finding 3: When positioning the device, ensuring good visibility of the targeted joints improves the accuracy of the results.

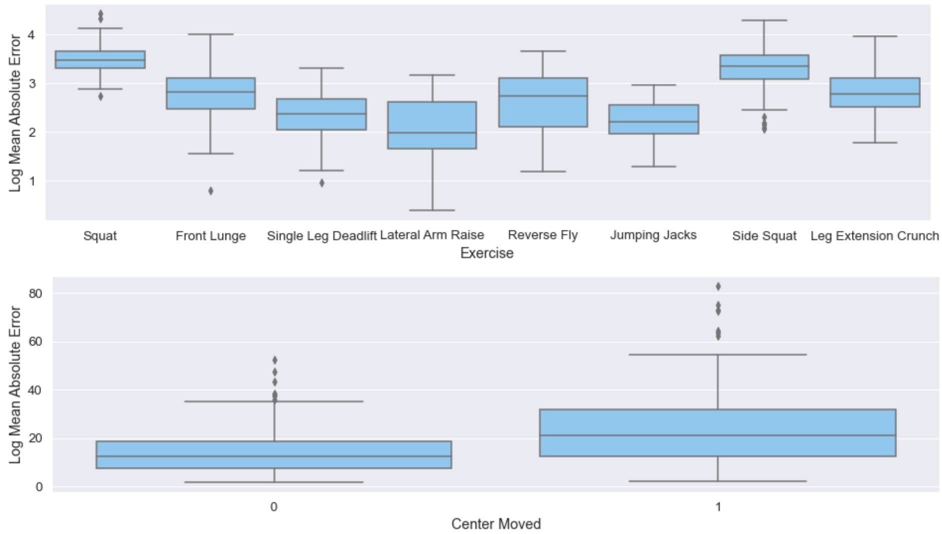


Figure 10. Boxplots representing the MAE in degrees on the logarithmic scale across all performed exercises and the *pelvic center moved* variable in the experiments. Both boxplots show significant differences in the mean and variance across the variables.

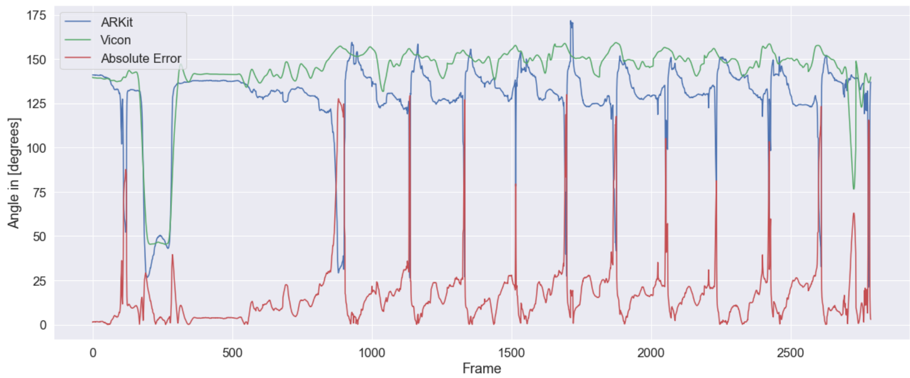


Figure 11. Left elbow angle of one of the subjects in the Single Leg Deadlift exercise, which shows several unexpected spikes during the execution. The spikes originate from ARKit incorrectly detecting the joint's position, most probably because of bad visibility of the elbow joint during the exercise.

5. Discussion

5.1. Factors Influencing ARKit's Performance

Based on the findings presented in Section 4, we identified several factors that influence the accuracy of ARKit's motion capture. The main requirement for good tracking is ensuring that the joints of interest are well visible to the camera and not hidden by other parts of the body during the movement. The exercise or motion itself is also of relevance. The results of the *t*-test hinted at a relevance of the coordinate system's stability during the exercise. However, this was not supported by the results of the logistic regression, so that the interpretation is unclear and requires further investigation.

The results of capturing human motion using ARKit could be influenced by several other factors, which were not further investigated within this research. This includes technical factors such as the device's processing power and additional sensors to improve the motion capture, the tracking environment such as lighting conditions or the background, or factors regarding the captured person, such as their clothing, body mass index, or skin color.

5.2. Bias of the Motion Capture Results

The upper body angles exposed a tendency of underestimation, and the results of the hips hinted at systematic overestimation as described in Section 3.1.2. Several values were located close to −1 or 1, which hints at a tendency to either systematic rather than cyclically occurring over- or underestimation. When aggregating the values for the different joints (Table 8), the results suggest that the upper body angles are underestimated, while the hip gets overestimated. The knee angles remain inconclusive with values relatively close to zero. They could hint at the mentioned cyclically occurring over- and underestimations or over- and underestimation based on the executed movement.

Table 8. The mean values of the ratio ME/MAE for the different joint angles.

Angle	Ratio ME/MAE
leftElbow	−0.46
rightElbow	−0.30
leftShoulder	−0.47
rightShoulder	−0.31
leftHip	0.59
rightHip	0.75
leftKnee	−0.19
rightKnee	0.01

5.3. Influence of the Tracked Joint Angle

The logistic regression results indicated a small, but significant effect of the lower body variable. These impressions are supported when inspecting the boxplot of the angles in the ME (Figure 12). The boxplot shows a tendency of underestimating the upper body angles, overestimating the hip angles, and a difference in the mean between the knee and hip angles. To investigate this effect, we performed the ANOVA analysis on the ME. We shifted the ME to only include positive values and applied the logarithmic transformation similar to our proceedings of the MAE as described in Section 2.6. The observed angles show an influence on the result ($\eta^2 = 0.26$, $p = 0.00$). Post-hoc analysis using Games-Howell supports the suggestions that the differences lie between the upper body angles and lower body angles and between the hip and knee angles (Appendix C, Table A2).

Interestingly, the exercise and movement of the hip center were the influencing factors for the MAE in contrast to the results of the ME. In the MAE, the difference between the angles is not observable anymore. The upper body error is mapped to a similar MAE as the lower body joints by only considering the absolute error (Figure 12). The ME for the whole dataset is −0.83°, meaning that overestimating the lower body joints and underestimating the upper body joints could be subject to error cancellation when considering the entire body. This effect could explain the MAE's dependency on the selected exercise while no dependency on the angle was observed.

The ANOVA results show an effect for the upper body variable and support the respective tendency of over-and underestimation. However, as explained in Section 2.6, the ME is prone to error cancellation effects. This unclear influence impacts the explanatory power, so we did not include these thoughts in the results and findings.

5.4. Impact of Incorrect Hip Detection

A commonly observed issue with the ARKit data were a reduced amplitude, and a baseline drift along the *y*-axis (see Figure 7), though the motion was tracked quite reliable. This issue was particularly the case for the lower body joints and led to a higher wMAE in those joints, but was also observed in other joints. In the screencasts of the recording, we often noticed that the detection of the hip joints was incorrect (Figure 13) and even varied during the execution of the exercise. Such shifts on the sagittal plane explain both the baseline drift and the amplitude reduction in the hip, knee, and shoulder angles, as all of them rely on the hip joints for their calculation. Especially from a side perspective, the hip joints allow for the most considerable deviations along the sagittal plane due to the amount or muscle and fat tissue around the pelvis. In the example of Figure 13, another issue aggravates the correct detection of the hip joints: the camera perspective was optimized for tracking the legs' position, which in this case means that the right joint hides the left hip joint. This positioning implies that ARKit needs to rely on other body landmarks to estimate its position. Finding an optimal camera position in which all joints are completely visible might not be possible for all movements.

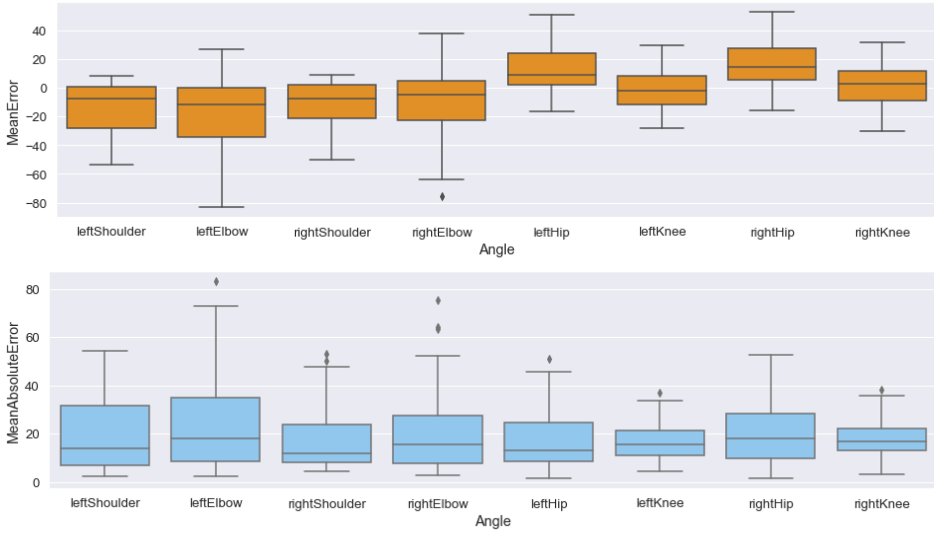


Figure 12. Boxplots representing the ME and MAE in degrees across all tracked angles in the experiments. The boxplots for the ME show a significant difference in the means of the upper and lower body angles, which is not visible for the MAE.

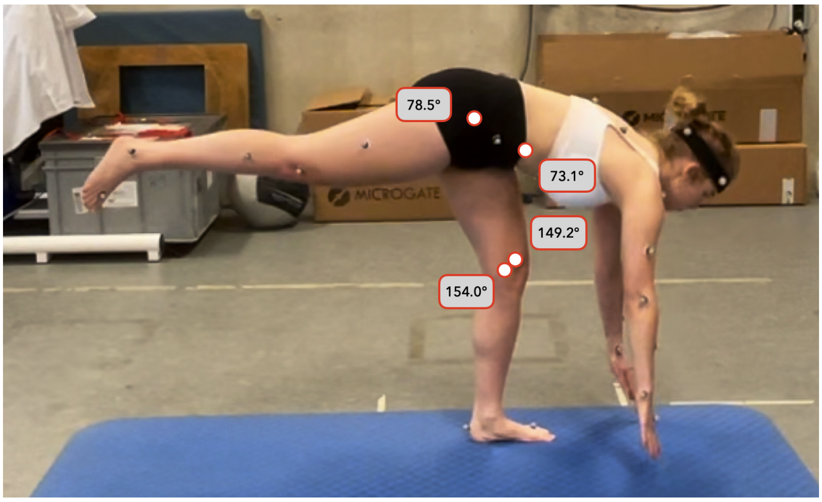


Figure 13. Exemplary screenshot of the frontal ARKit recording of one subject during the Single Leg Deadlift exercise, showing a bad detection of the hip joints and confusion of the knee joints.

5.5. Improving the ARKit Data during Post-Processing

The good correlation results opened up the question of whether it is possible to improve the ARKit motion capture data through post-processing to approximate the Vicon data. A systematic error concerning detecting the hip joints in a position too far anterior is a possible explanation and is subject to further investigation. If this is the case, both the baseline shift and the amplitude reduction could be corrected by applying a scale factor and shifting the data on the y-axis. Compensating the baseline shift would reduce the wMAE results by 7.61° and lead to more reliable and accurate results. However, no systematic error could be found when shifting the ARKit data along the y-axis by vertically shifting the ARKit data (Figure 9). The observed shift instead seems to be caused by other factors such as the incorrect detection of joints.

During the data analysis, we used a sliding window approach to maximizing the cross-correlation between the ARKit and Vicon data to compensate for possible time lags, as no synchronization of the iPads and the Vicon system was possible during the experiment. Possible reasons for lags are different hardware clocks and the delay of the body detection algorithm of the ARKit framework. The sliding window was set to a maximum of 120 frames, which equals approximately 2 s, only to allow reasonable shifts within the exercises and compensate for the lag caused by technical limitations. The approach was chosen to maximize the comparability between the results of the two systems. However, as the sliding window approach was applied individually to each angle, exercise, subject, and view, each configuration was shifted to its optimal result within the given time window. This approach does not consider potential lags within ARKit's motion capture, for example, a slower recognition of changes for some parts of the recognized body.

5.6. Comparing the Results of 2D and 3D Motion Capture Systems

As stated in the analysis of Sarafianos et al. [31], monocular video-based motion capture systems exhibit several limitations, which reduce their applicability to real-world scenarios. Among the most significant limitations are the ambiguities of the detected poses due to occlusion and distortion of the camera image caused by the camera's viewing angle and position [31], which is a relevant limitation in both 2D and 3D motion capture systems. In this research, we were able to show that ARKit, as an example for 3D motion capture systems supported by different smartphone sensors, is robust against a variation of 30° regarding the positioning of the device. The factor analysis did not expose an influence of the device position. However, poor visibility of joints still led to significant decreases in the accuracy of the measured angles. Mobile 3D motion capture frameworks based on monocular video data such as ARKit improve some of the limitations of 2D motion capture systems but cannot overcome them completely.

5.7. Potential Use Cases for Mobile 3D Motion Capture-Based Applications

The findings of this research raise the question of possible application areas for human motion capture using mobile 3D motion capture frameworks such as Apple ARKit. Referring to the three categories defined by Moeslund et al. [2], such frameworks could be applied to use cases in categories (2) interacting with software or (3) motion analysis for medical examinations or performance analysis, as it focuses on tracking single bodies rather than observing crowds. The results suggest that ARKit can track a motion's progression reliably but with relatively high error rates, depending on the joint of interest. Human motion capture using ARKit is further limited to a relatively small set of trackable joints. For example, the hand and toe joints are not actively tracked but calculated based on the angle and wrist joints, limiting the trackable joint angles to the shoulder, elbow, hip, and knee. However, mobile 3D motion capture frameworks are a promising technology for use cases that focus on tracking a specific motion of body parts rather than the exact joint position. Such use cases can be seen in category (2), such as interacting with software through gestures or other movements. Potential use cases in (3) include sports applications for amateurs or physiotherapy applications, which could focus on counting repetitions of a specific exercise. Depending on the motion and joint of interest, specific use cases relying on the exact joint position and angle data might be possible if the two main requirements for a good tracking presented at the beginning of this section can be met. For example, such use cases could include measuring the possible range of motion of a joint before and after a particular intervention and monitoring the progress in the medical field, or correcting the execution of a specific exercise in sports and physiotherapy applications. Using mobile 3D motion capture frameworks in these use cases would extend the usage of human motion capture technologies beyond professional settings and allow day-to-day usage at home, performed by consumers. ARKit and other mobile IPS systems enable new use cases, especially in mHealth, which were not possible with previous HMC systems. Our findings show how mobile 3D motion capture frameworks can be applied and how mHealth applications could leverage the software for future applications. However, the limitations of 3D motion capture frameworks and ARKit's boundaries, in particular, need to be considered and should be evaluated before applying the technology to specific use cases.

5.8. Limitations

The design of this research includes several limitations. While the lab experiment produced a data set of over 1000 exercise executions, the data were collected from ten study participants only due to the restrictions caused by the ongoing COVID-19 pandemic. The limited number of participants might limit the *external validity* of this research. The participants' traits further limit the external validity. While covering heights between 156 cm and 198 cm, their body mass index was in a normal range. In addition, all participants had a lighter skin tone. The experiment was conducted in a laboratory with controlled background and lighting conditions.

Even though the study setup aimed at reducing possible influences on the study's *internal validity* which were not part of the observation, the impact of additional factors cannot be eliminated. Possible factors include the influence of the specific performance of the exercises by the subjects or the effect of the clothing worn. Furthermore, the subjects were recruited from the social surroundings of the researchers. They might not be representative of the whole population. The internal validity is further affected by the sliding window approach to compensate for the time lag due to missing clock synchronization and processing time. While the approach is limited to a maximum window of approximately two seconds, this shift could still have improved the results above the observable results. Additionally, the data set contained a reduced amount of exercise data for the upper body joints due to the export problems of the iPad on the side position. We applied the Welch ANOVA test to identify dependencies of the MAE instead of the ANOVA test, as the variance of the individual factors was not equally distributed. However, another prerequisite for (Welch) ANOVA and Welch *t*-test, normally distributed data, was only partially given for the MAE, even though the ANOVA analysis is said to be quite robust against this problem. We applied a logarithmic transformation to the data before performing the ANOVA and *t*-tests to overcome these limitations. Moreover, the observations used in (Welch) ANOVA should be independent of each other. In our experiment setup, the recording of angle motion happened simultaneously in all subjects and exercises. The observed angle deviations of the systems are expected to be independent. However, a poorly tracked angle might cause a higher risk to affect another angle's accuracy in a real-world scenario. Thus, the assumption of independent observations is hard to verify. Moreover, ARKit is only one example of a mobile 3D motion capture framework. Other frameworks rely on different technologies and algorithms and could exhibit different results and limitations.

6. Conclusions

This research evaluated mobile 3D motion capture based on the example of ARKit, Apple's framework for smartphone-based 3D motion capture. In contrast to existing monocular motion capture software, ARKit detects the human body in a 3-dimensional space instead of only two dimensions and augments its results by using smartphone sensor data such as IMU or depth data from the integrated LiDAR sensor. Our laboratory experiment, including ten participants, investigated ARKit's accuracy and influencing factors in eight body-weight exercises and compared it to the Vicon system, a gold standard for human motion capture. Our results provide evidence that mobile 3D motion capture frameworks can track the motion's progression with reasonable accuracy but with relatively high mean absolute error rates. The accuracy mainly depends on two factors: the visibility of the joints of interest and the observed motion. In contrast to 2D systems, the 3D motion capture framework exposed certain robustness against the positioning of the camera. However, similar limitations regarding the tracking of poorly visible joints remain.

Mobile 3D motion capture frameworks are promising and lightweight mobile technologies which could enable new use cases for human-computer interaction through motion or application in health and medical fields. Their limitations, especially regarding the relatively high error rates compared to the gold standard system, need to be considered for each use case.

Author Contributions: Conceptualization, L.M.R., M.K., T.F. and S.M.J.; methodology, L.M.R., M.K. and S.M.J.; software, L.M.R.; validation, L.M.R., M.K., T.F. and S.M.J.; formal analysis, M.K. and L.M.R.; investigation, L.M.R. and T.F.; resources, L.M.R. and S.M.J.; data curation, M.K.; writing—original draft preparation, L.M.R., M.K. and T.F.; writing—review and editing, L.M.R., M.K. and S.M.J.; visualization, L.M.R. and M.K.; supervision, L.M.R. and S.M.J.; project administration, L.M.R.; funding acquisition, L.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from Software Campus through the German Federal Ministry of Education and Research, grant number 01IS17049.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Technical University of Munich (Proposal 515/21 S on 19 August 2021). All participants were informed about the aims of the study and gave their consent about the publication of the anonymized data.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the subjects to publish this paper.

Data Availability Statement: All data is available on Zenodo [49].

Acknowledgments: We want to thank Florian Kreuzpointner for his support during the planning and execution of the study as well as the participants of the study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AOI	Angles of Interest
EMS	Electromagnetig Measurement Systems
FL	Front Lunge
HMC	Human Motion Capture
IPS	Image Processing Systems
IMU	Inertial Measurement Unit
JJ	Jumping Jacks
LAR	Lateral Arm Raise
LEC	Leg Extension Crunch
LE	Left Elbow
LH	Left Hip
LK	Left Knee
LS	Left Shoulder
MAE	Mean Absolute Error
ME	Mean Error
OMS	Optoelectronic Measurement Systems
PCC	Pearson Correlation Coefficient
RE	Right Elbow
RF	Reverse Fly
RH	Right Hip
RK	Right Knee
RS	Right Shoulder
S	Squat
SDK	Software Development Kit
SS	Side Squat

SLD Single Leg Deadlift
ULS Ultrasonic Localization Systems
wMAE Weighted Mean Absolute Error

Appendix A. Distributions of the Factors Used in the Welch ANOVA Analysis

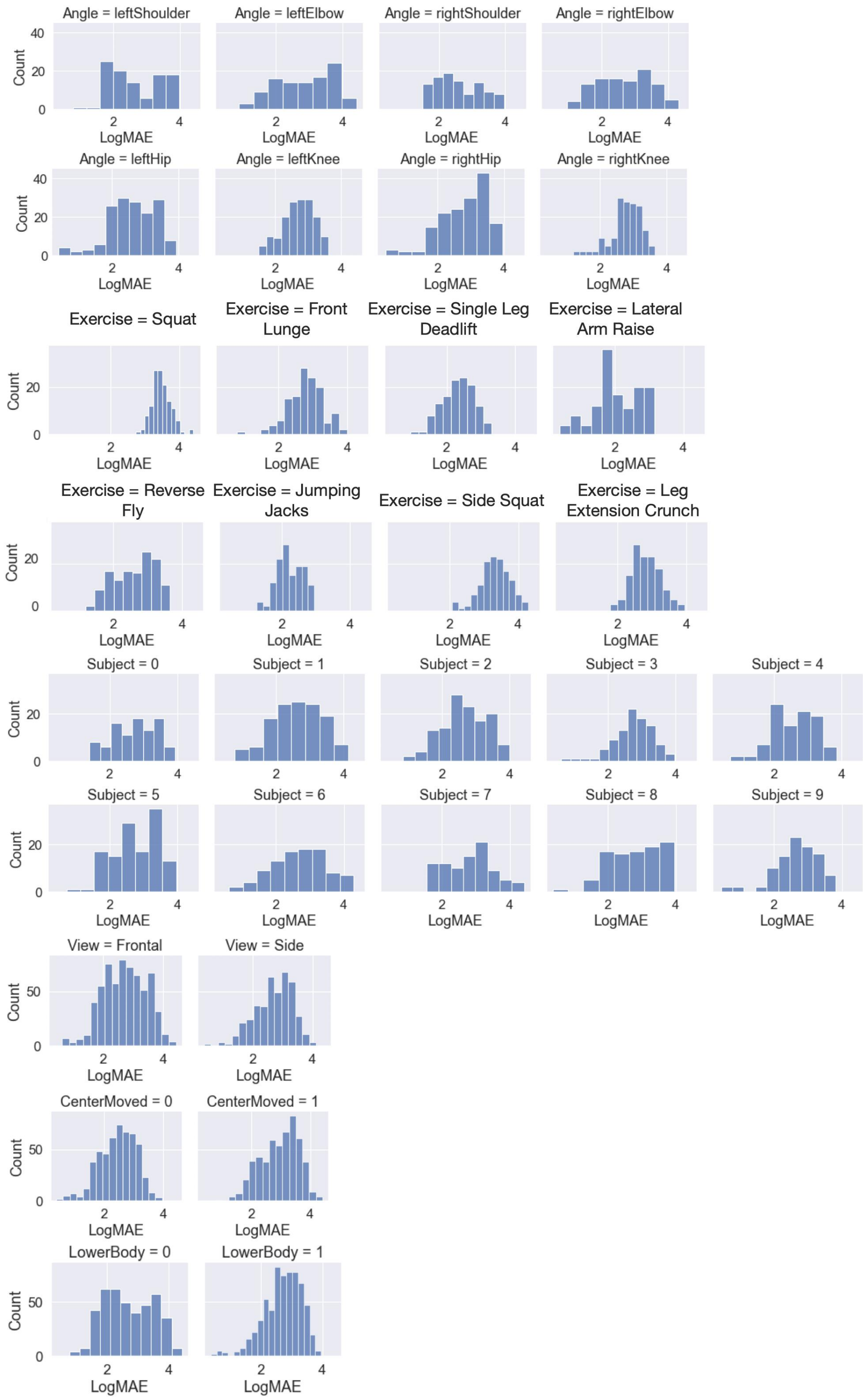


Figure A1. Distributions of the individual factors of the MAE on the logarithmic scale used in the factor analysis. Due to the transformation on the logarithmic scale, all factors are sufficiently close to a normal distribution, so that a factor analysis using Welch ANOVA/*t*-tests should be possible.

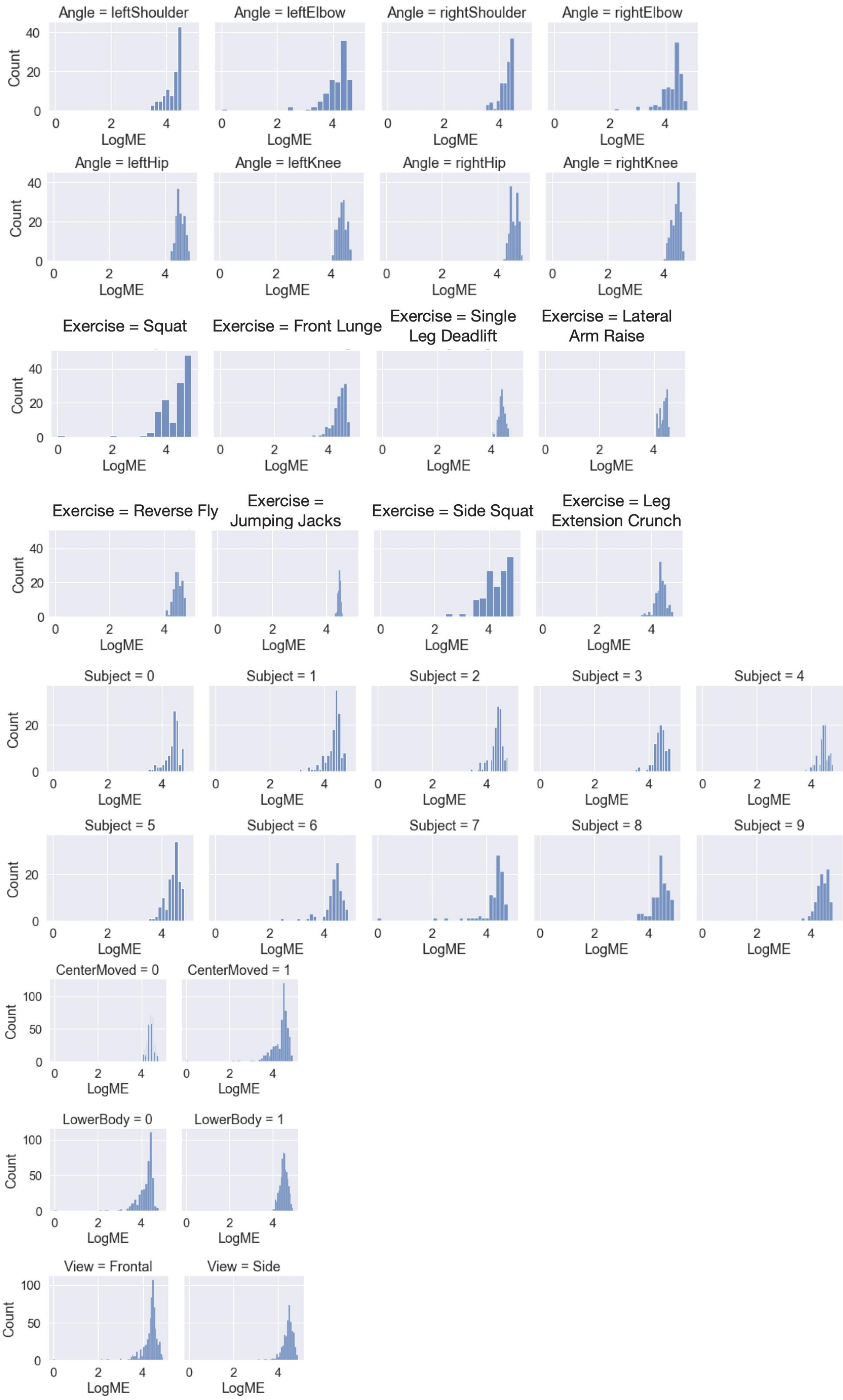


Figure A2. Distributions of the individual factors of the ME on the logarithmic scale used in the Welch ANOVA analysis. All of the factors show a distribution which is sufficiently close to a normal distribution so that an ANOVA analysis should be possible.

Appendix B. Bias

Exercise	Angle / View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	-27.98 ± 17.23	-20.72 ± 10.15	0.75 ± 3.59	-8.23 ± 5.2	-0.42 ± 8.77	19.84 ± 7.1	3.99 ± 8.63	9.04 ± 5.61
	Side	-28.08 ± 12.71	-30.91 ± 3.41	-10.38 ± 2.68	-14.8 ± 1.8	-3.52 ± 5.12	22.17 ± 3.25	10.16 ± 6.7	13.7 ± 4.29
Jumping Jacks	Frontal	2.58 ± 4.94	3.19 ± 3.71	2.35 ± 2.03	2.79 ± 2.28	6.29 ± 2.86	7.13 ± 2.6	-1.08 ± 3.85	0.45 ± 4.0
	Side	8.39 ± 1.52	6.48 ± 1.49	0.5 ± 2.51	5.6 ± 0.76	7.17 ± 3.32	9.58 ± 2.84	1.79 ± 4.62	2.77 ± 4.27
Lateral Arm Raise	Frontal	-5.46 ± 5.62	-6.03 ± 3.76	1.26 ± 3.42	3.04 ± 3.72	3.9 ± 4.11	3.88 ± 3.73	-17.53 ± 4.21	-17.57 ± 3.29
	Side	-2.48 ± 1.92	-3.48 ± 2.27	-0.88 ± 2.38	3.24 ± 1.98	1.59 ± 4.95	6.62 ± 3.62	-17.69 ± 4.48	-17.33 ± 3.83
Leg Extension Crunch	Frontal	-0.22 ± 11.99	16.22 ± 7.8	-28.58 ± 9.15	-14.62 ± 5.89	-3.52 ± 3.39	4.17 ± 4.25	-9.32 ± 4.61	-6.28 ± 5.0
	Side	9.42 ± 16.08	24.91 ± 12.09	-35.27 ± 3.82	-19.19 ± 6.97	-4.8 ± 5.41	8.31 ± 5.88	-13.31 ± 4.26	-15.84 ± 7.75
Reverse Fly	Frontal	-2.83 ± 7.09	3.87 ± 5.78	-5.45 ± 6.48	-3.01 ± 5.7	24.94 ± 6.35	27.43 ± 6.75	5.26 ± 12.58	5.78 ± 12.27
	Side	-0.65 ± 5.28	4.61 ± 2.58	-11.51 ± 2.06	-2.23 ± 4.3	22.11 ± 7.19	20.94 ± 7.54	6.77 ± 12.56	3.64 ± 10.85
Side Squat	Frontal	-46.02 ± 14.64	-41.15 ± 15.21	-38.13 ± 11.12	-33.23 ± 9.59	22.55 ± 7.62	36.39 ± 5.22	-12.39 ± 9.89	15.85 ± 6.24
	Side	-46.08 ± 5.2	-27.6 ± 4.03	-28.62 ± 3.56	-20.89 ± 1.9	26.08 ± 8.27	30.22 ± 3.97	-11.36 ± 10.51	11.55 ± 4.1
Single Leg Deadlift	Frontal	-18.27 ± 9.06	-3.3 ± 8.26	-3.24 ± 5.66	-2.05 ± 5.62	9.28 ± 3.17	-3.04 ± 6.62	-3.36 ± 3.67	-4.29 ± 7.56
	Side	-12.78 ± 0.73	-4.96 ± 1.74	-1.72 ± 5.84	2.67 ± 4.01	11.4 ± 5.31	-4.16 ± 7.3	-4.21 ± 4.01	-5.72 ± 7.36
Squat	Frontal	-43.14 ± 18.84	-36.95 ± 17.32	-34.59 ± 11.14	-32.9 ± 10.54	34.96 ± 6.65	37.05 ± 6.75	14.72 ± 6.47	15.85 ± 7.71
	Side	-44.0 ± 11.41	-31.72 ± 9.74	-30.3 ± 6.29	-23.41 ± 2.55	31.45 ± 5.14	29.88 ± 4.98	14.16 ± 5.0	9.89 ± 6.24

Figure A3. Pivot Table of the average Mean Error (ME) distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker purple color hinting at underestimation and darker orange color hinting at overestimation. Values closer to zero either indicate good performance or error cancellation.

Exercise	Angle / View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	-0.79	-0.87	0.09	-0.67	-0.04	0.98	0.29	0.65
	Side	-0.77	-0.92	-0.71	-0.84	-0.28	1.00	0.58	0.76
Jumping Jacks	Frontal	0.34	0.45	0.36	0.37	0.75	0.86	-0.07	0.03
	Side	0.92	0.70	0.08	0.72	0.76	0.96	0.12	0.20
Lateral Arm Raise	Frontal	-0.73	-0.82	0.19	0.45	0.78	0.81	-0.99	-1.00
	Side	-0.66	-0.64	-0.16	0.57	0.32	0.97	-1.00	-0.99
Leg Extension Crunch	Frontal	-0.01	0.83	-0.89	-0.84	-0.32	0.36	-0.63	-0.38
	Side	0.48	0.96	-0.97	-0.90	-0.34	0.51	-0.78	-0.79
Reverse Fly	Frontal	-0.32	0.38	-0.46	-0.28	0.99	0.99	0.34	0.39
	Side	-0.08	0.31	-0.78	-0.26	0.99	1.00	0.44	0.30
Side Squat	Frontal	-0.98	-0.97	-0.91	-0.92	1.00	1.00	-0.75	0.60
	Side	-0.98	-0.99	-0.95	-0.88	1.00	1.00	-0.70	0.54
Single Leg Deadlift	Frontal	-0.87	-0.38	-0.28	-0.19	0.88	-0.21	-0.39	-0.46
	Side	-0.96	-0.81	-0.18	0.37	0.93	-0.28	-0.52	-0.57
Squat	Frontal	-0.97	-0.98	-0.88	-0.88	0.99	0.99	0.49	0.52
	Side	-0.97	-1.00	-0.83	-0.84	0.98	0.99	0.51	0.42

Figure A4. Pivot Table of the ratio of the ME divided by the MAE distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise. Values close to zero indicate either good performance of the tracking or over- and underestimation canceling each other out. Values closer to -1 and 1 hint at systematic under- and overestimation in the specific configuration.

Appendix C. ANOVA Post-Hoc Analysis

Appendix C.1. Mean Absolute Error

Table A1. The results of the ANOVA Post-hoc analysis of the MAE for the eight exercises Front Lunge (FL), Jumping Jacks (JJ), Lateral Arm Raise (LAR), Leg Extension Crunch (LEC), Reverse Fly (RF), Side Squat (SS), Single Leg Deadlift (SLD), and Squat (S).

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
FL	JJ	2.78	2.25	0.53	0.06	9.69	242.95	0.00	0.26
FL	LAR	2.78	2.04	0.74	0.07	10.05	240.85	0.00	0.28
FL	LEC	2.78	2.81	−0.03	0.06	−0.49	254.87	1.00	0.00
FL	RF	2.78	2.61	0.17	0.07	2.53	254.89	0.19	0.02
FL	SS	2.78	3.32	−0.54	0.06	−9.35	257.60	0.00	0.25
FL	SLD	2.78	2.33	0.45	0.06	7.51	253.76	0.00	0.18
FL	S	2.78	3.49	−0.71	0.05	−14.17	204.84	0.00	0.43
JJ	LAR	2.25	2.04	0.21	0.07	3.12	204.18	0.04	0.04
JJ	LEC	2.25	2.81	−0.56	0.05	−11.29	258.38	0.00	0.33
JJ	RF	2.25	2.61	−0.36	0.06	−5.85	221.58	0.00	0.12
JJ	SS	2.25	3.32	−1.07	0.05	−21.28	255.86	0.00	0.63
JJ	SLD	2.25	2.33	−0.08	0.05	−1.51	238.68	0.80	0.01
JJ	S	2.25	3.49	−1.24	0.04	−30.34	241.51	0.00	0.78
LAR	LEC	2.04	2.81	−0.77	0.07	−11.00	219.24	0.00	0.31
LAR	RF	2.04	2.61	−0.57	0.08	−7.24	236.94	0.00	0.17
LAR	SS	2.04	3.32	−1.28	0.07	−18.19	224.12	0.00	0.56
LAR	SLD	2.04	2.33	−0.29	0.07	−4.03	230.32	0.00	0.06
LAR	S	2.04	3.49	−1.45	0.06	−22.61	173.71	0.00	0.66
LEC	RF	2.81	2.61	0.20	0.06	3.13	236.94	0.04	0.04
LEC	SS	2.81	3.32	−0.52	0.05	−9.96	261.64	0.00	0.26
LEC	SLD	2.81	2.33	0.48	0.06	8.66	249.27	0.00	0.23
LEC	S	2.81	3.49	−0.68	0.04	−15.40	226.49	0.00	0.47
RF	SS	2.61	3.32	−0.71	0.06	−11.10	241.49	0.00	0.32
RF	SLD	2.61	2.33	0.28	0.07	4.22	244.66	0.00	0.07
RF	S	2.61	3.49	−0.88	0.06	−15.39	186.05	0.00	0.47
SS	SLD	3.32	2.33	0.99	0.06	17.69	251.45	0.00	0.55
SS	S	3.32	3.49	−0.17	0.04	−3.63	221.60	0.01	0.05
SLD	S	2.33	3.49	−1.16	0.05	−24.28	200.97	0.00	0.70

Appendix C.2. Mean Error

Table A2. The results of the ANOVA Post-hoc analysis of the ME for the eight angles left elbow (LE), left hip (LH), left knee (LK), left shoulder (LS), right elbow (RE), right hip (RH), right knee (RK), and right shoulder (RS).

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
LE	LH	4.10	4.55	−0.45	0.06	−7–73	110.78	0.00	0.19
LE	LK	4.10	4.40	−0.29	0.06	−5.03	111.83	0.00	0.09
LE	LS	4.10	4.22	−0.11	0.06	−1.78	148.18	0.63	0.01
LE	RE	4.10	4.25	−0.15	0.07	−2.14	177.59	0.39	0.02
LE	RH	4.10	4.60	−0.50	0.06	−8.54	110.20	0.00	0.23
LE	RK	4.10	4.44	−0.34	0.06	−5.74	111.85	0.00	0.12
LE	RS	4.10	4.27	−0.17	0.06	−2.69	134.33	0.14	0.03
LH	LK	4.55	4.40	0.16	0.02	9.12	315.23	0.00	0.21
LH	LS	4.55	4.22	0.34	0.03	11.13	138.85	0.00	0.33
LH	RE	4.55	4.25	0.30	0.04	7.63	121.90	0.00	0.19
LH	RH	4.55	4.60	−0.05	0.02	−2.81	313.63	0.10	0.02
LH	RK	4.55	4.44	0.12	0.02	6.71	315.19	0.00	0.12
LH	RS	4.55	4.27	0.28	0.03	11.03	155.72	0.00	0.33
LK	LS	4.40	4.22	0.18	0.03	5.92	143.19	0.00	0.12
LK	RE	4.40	4.25	0.15	0.04	3.68	124.27	0.01	0.05

Table A2. *Cont.*

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
LK	RH	4.40	4.60	−0.20	0.02	−11.99	313.81	0.00	0.31
LK	RK	4.40	4.44	−0.04	0.02	−2.34	318.00	0.28	0.02
LK	RS	4.40	4.27	0.13	0.03	4.91	161.84	0.00	0.09
LS	RE	4.22	4.25	−0.03	0.05	−0.72	187.27	1.00	0.00
LS	RH	4.22	4.60	−0.38	0.03	−12.72	136.43	0.00	0.39
LS	RK	4.22	4.44	−0.22	0.03	−7.26	143.28	0.00	0.17
LS	RS	4.22	4.27	−0.05	0.04	−1.45	196.84	0.83	0.01
RE	RH	4.25	4.60	−0.35	0.04	−8.82	120.58	0.00	0.24
RE	RK	4.25	4.44	−0.19	0.04	−4.71	124.32	0.00	0.08
RE	RS	4.25	4.27	−0.02	0.04	−0.41	167.96	1.00	0.00
RH	RK	4.60	4.44	0.16	0.02	9.54	313.75	0.00	0.22
RH	RS	4.60	4.27	0.33	0.03	12.90	152.29	0.00	0.40
RK	RS	4.44	4.27	0.17	0.03	6.49	161.97	0.00	0.14

References

1. Moeslund, T.B.; Granum, E. A Survey of Computer Vision-Based Human Motion Capture. *Comput. Vis. Image Underst.* **2001**, *81*, 231–268. [CrossRef]
2. Moeslund, T.B.; Hilton, A.; Krüger, V. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **2006**, *104*, 90–126. [CrossRef]
3. Chiari, L.; Croce, U.D.; Leardini, A.; Cappozzo, A. Human movement analysis using stereophotogrammetry. *Gait Posture* **2005**, *21*, 197–211. [CrossRef] [PubMed]
4. Elliott, B.; Alderson, J. Laboratory versus field testing in cricket bowling: A review of current and past practice in modelling techniques. *Sports Biomech.* **2007**, *6*, 99–108. [CrossRef] [PubMed]
5. Carse, B.; Meadows, B.; Bowers, R.; Rowe, P. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3D motion analysis system. *Physiotherapy* **2013**, *99*, 347–351. [CrossRef] [PubMed]
6. McLean, S.G. Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *Br. J. Sports Med.* **2005**, *39*, 355–362. [CrossRef]
7. van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**, *18*, 806–819. [CrossRef]
8. Belyea, B.C.; Lewis, E.; Gabor, Z.; Jackson, J.; King, D.L. Validity and Intrarater Reliability of a 2-Dimensional Motion Analysis Using a Handheld Tablet Compared With Traditional 3-Dimensional Motion Analysis. *J. Sport Rehabil.* **2015**, *24*, 2014-0194. [CrossRef]
9. Paul, S.S.; Lester, M.E.; Foreman, K.B.; Dibble, L.E. Validity and Reliability of Two-Dimensional Motion Analysis for Quantifying Postural Deficits in Adults With and Without Neurological Impairment. *Anat. Rec.* **2016**, *299*, 1165–1173. [CrossRef]
10. Springer, S.; Seligmann, G.Y. Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors* **2016**, *16*, 194. [CrossRef]
11. Puh, U.; Hoehlein, B.; Deutsch, J.E. Validity and Reliability of the Kinect for Assessment of Standardized Transitional Movements and Balance. *Phys. Med. Rehabil. Clin. N. Am.* **2019**, *30*, 399–422. [CrossRef]
12. Schärer, C.; Siebenthal, L.V.; Lomax, I.; Gross, M.; Taube, W.; Hübner, K. Simple Assessment of Height and Length of Flight in Complex Gymnastic Skills: Validity and Reliability of a Two-Dimensional Video Analysis Method. *Appl. Sci.* **2019**, *9*, 3975. [CrossRef]
13. Alahmari, A.; Herrington, L.; Jones, R. Concurrent validity of two-dimensional video analysis of lower-extremity frontal plane of movement during multidirectional single-leg landing. *Phys. Ther. Sport* **2020**, *42*, 40–45. [CrossRef]
14. Vicon Motion Capture Systems. Available online: https://www.vicon.com (accessed on 26 January 2022).
15. Qualisys Motion Capture Systems. Available online: https://www.qualisys.com (accessed on 26 January 2022).
16. Xsens Motion Capture Systems. Available online: https://www.xsens.com (accessed on 26 January 2022).
17. Perception Neuron Motion Capture. Available online: https://neuronmocap.com/ (accessed on 26 January 2022).
18. Stelzer, A.; Pourvoyeur, K.; Fischer, A. Concept and Application of LPM—A Novel 3-D Local Position Measurement System. *IEEE Trans. Microw. Theory Tech.* **2004**, *52*, 2664–2669. [CrossRef]
19. OpenPose: Real-Time Multi-Person Keypoint Detection Library for Body, Face, Hands, and Foot Estimation. Available online: https://github.com/CMU-Perceptual-Computing-Lab/openpose (accessed on 26 January 2022).
20. ARKit: Capturing Body Motion in 3D. Available online: https://developer.apple.com/documentation/arkit/content_anchors/capturing_body_motion_in_3d (accessed on 26 January 2022).
21. Vision: Detecting Human Body Poses in Images. Available online: https://developer.apple.com/documentation/vision/detecting_human_body_poses_in_images (accessed on 26 January 2022).
22. TensorFlow Pose Estimate. Available online: https://www.tensorflow.org/lite/examples/pose_estimation/overview (accessed on 26 January 2022).

23. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [CrossRef] [PubMed]
24. Whittle, M.W. Clinical gait analysis: A review. *Hum. Mov. Sci.* **1996**, *15*, 369–387. [CrossRef]
25. Oyebode, O.; Ndulue, C.; Alhasani, M.; Orji, R. Persuasive Mobile Apps for Health and Wellness: A Comparative Systematic Review. In *Lecture Notes in Computer Science*; Springer International Publishing: Zurich, Switwerland, 2020; pp. 163–181. [CrossRef]
26. research2guidance. Number of Downloads of mHealth Apps Worldwide from 2013 to 2018 (in Billions) [Graph]. 2018. Available online: https://de-statista-com/statistik/daten/studie/695434/umfrage/nummer-der-weltweiten-downloads-von-mhealth-apps/ (accessed on 26 January 2022).
27. Schoeppe, S.; Alley, S.; Lippevelde, W.V.; Bray, N.A.; Williams, S.L.; Duncan, M.J.; Vandelanotte, C. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2016**, *13*, 127. [CrossRef]
28. Boulos, M.N.K.; Brewer, A.C.; Karimkhani, C.; Buller, D.B.; Dellavalle, R.P. Mobile medical and health apps: State of the art, concerns, regulatory control and certification. *Online J. Public Health Inform.* **2014**, *5*, 229. [CrossRef]
29. Lopes, T.J.A.; Ferrari, D.; Ioannidis, J.; Simic, M.; Azevedo, F.M.D.; Pappas, E. Reliability and Validity of Frontal Plane Kinematics of the Trunk and Lower Extremity Measured with 2-Dimensional Cameras During Athletic Tasks: A Systematic Review with Meta-analysis. *J. Orthop. Sports Phys. Ther.* **2018**, *48*, 812–822. [CrossRef]
30. Zago, M.; Luzzago, M.; Marangoni, T.; Cecco, M.D.; Tarabini, M.; Galli, M. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Front. Bioeng. Biotechnol.* **2020**, *8*, 181. [CrossRef]
31. Sarafianos, N.; Boteanu, B.; Ionescu, B.; Kakadiaris, I.A. 3D Human pose estimation: A review of the literature and analysis of covariates. *Comput. Vis. Image Underst.* **2016**, *152*, 1–20. [CrossRef]
32. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. *Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*; CVPR: Prague, Czech Republic, 2017.
33. Simon, T.; Joo, H.; Matthews, I.; Sheikh, Y. *Hand Keypoint Detection in Single Images Using Multiview Bootstrapping*; CVPR: Prague, Czech Republic, 2017.
34. Wei, S.E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. *Convolutional Pose Machines*; CVPR: Prague, Czech Republic, 2016. [CrossRef]
35. D'Antonio, E.; Taborri, J.; Palermo, E.; Rossi, S.; Patane, F. A markerless system for gait analysis based on OpenPose library. In Proceedings of the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 25–28 May 2020; IEEE: Dubrovnik, Croatia, 2020. [CrossRef]
36. Ota, M.; Tateuchi, H.; Hashiguchi, T.; Kato, T.; Ogino, Y.; Yamagata, M.; Ichihashi, N. Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait Posture* **2020**, *80*, 62–67. [CrossRef] [PubMed]
37. Nakano, N.; Sakura, T.; Ueda, K.; Omura, L.; Kimura, A.; Iino, Y.; Fukashiro, S.; Yoshioka, S. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Front. Sports Act. Living* **2020**, *2*, 50. [CrossRef] [PubMed]
38. MediaPipe Pose. Available online: https://google.github.io/mediapipe/solutions/pose.html (accessed on 18 March 2022).
39. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. BlazePose: On-device Real-time Body Pose tracking. *arXiv* **2020**, arXiv:2006.10204.
40. Zhou, X.; Leonardos, S.; Hu, X.; Daniilidis, K. 3D shape estimation from 2D landmarks: A convex relaxation approach. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–15 June 2015; IEEE: Boston, MA, USA, 2015; pp. 4447–4455. [CrossRef]
41. Zhou, X.; Zhu, M.; Leonardos, S.; Derpanis, K.G.; Daniilidis, K. Sparseness Meets Deepness: 3D Human Pose Estimation from Monocular Video. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Las Vegas, NV, USA, 2016; pp. 4966–4975. [CrossRef]
42. Akhter, I.; Black, M.J. Pose-conditioned joint angle limits for 3D human pose reconstruction. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–15 June 2015; IEEE: Boston, MA, USA, 2015; pp. 1446–1455. [CrossRef]
43. Ma, F.; Cavalheiro, G.V.; Karaman, S. Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Montreal, QC, Canada, 2019; pp. 3288–3295. [CrossRef]
44. Reimer, L.M.; Weigel, S.; Ehrenstorfer, F.; Adikari, M.; Birkle, W.; Jonas, S. Mobile Motion Tracking for Disease Prevention and Rehabilitation Using Apple ARKit. In *Studies in Health Technology and Informatics*; Hayn, D., Schreier, G., Baumgartner, M., Eds.; IOS Press: Amsterdam, The Netherlands, 2021. [CrossRef]
45. Basiratzadeh, S.; Lemaire, E.D.; Baddour, N. Augmented Reality Approach for Marker-based Posture Measurement on Smartphones. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Montreal, QC, Canada, 2020; pp. 4612–4615. [CrossRef]
46. Full Body Modeling with Plug-In Gait. Available online: https://docs.vicon.com/display/Nexus212/Full+body+modeling+with+Plug-in+Gait (accessed on 26 January 2022).

47. Schmider, E.; Ziegler, M.; Danay, E.; Beyer, L.; Bühner, M. Is It Really Robust?: Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. *Methodology* **2010**, *6*, 147–151. [CrossRef]
48. Games, P.A.; Howell, J.F. Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *J. Educ. Stat.* **1976**, *1*, 113–125.
49. Reimer, L.M.; Kapsecker, M.; Fukushima, T.; Jonas, S.M. A Dataset for Evaluating 3D Motion Captured Synchronously by ARKit and Vicon. *ZENODO* **2022**. [CrossRef]

6 Publication C: mBalance: Detect Postural Imbalance with Mobile Devices

This publication presents one of the use cases of *ARKit* in the health sector: the detection of postural imbalance using optical motion tracking on mobile devices. It presents the implementation of an app prototype, which digitizes the Romberg test for imbalance detection and provides an evaluation of the app's detection accuracy and usability.

Conference: dHealth 2022
Number of Pages: 9
Type: Full Paper
Review: Peer Reviewed (2 Reviewers)

Summary

Problem. Postural imbalance is a symptom of several diseases, such as neurological diseases and diseases and injuries of the musculoskeletal system or the vestibular system. It increases the risk of falls and injuries and should be monitored and treated. The severity of postural imbalance is volatile and can change based on the daily conditions of the patient. Postural imbalance should be closely monitored, which is currently mostly done as part of regular examinations by a physician. Even though the patients can perform balance assessments in between assessments, the assessment quality is often subjective, or assessments are not performed at all. **Objective.** This research proposes *mBalance*, a mobile application that allows patients to perform a balance assessment using the established Romberg test at home. *mBalance* guides the patient through the assessment using audio commands and tracks the patient's balance state with optical motion tracking. It is an accessible and affordable monitoring tool that can generate a reliable data basis for patients and physicians. **Methods.** We describe the concept and implementation of a prototypical version of the *mBalance* application. We evaluated the prototype's accuracy in detecting imbalance and usability in two separate laboratory experiments with 30 and 31 healthy participants, respectively. We asked the subjects

to perform three subsequent balance assessments in the validation study. In the first two assessments, they should maintain their balance. In the third assessment, they should simulate a loss of balance by leaving the test position by opening up their arms or taking a step. The subjects were asked to perform a balance assessment using the Thinking Aloud method while filmed by a camera in the usability evaluation. The results were transcribed, and common usability problems were grouped into categories and analyzed. **Results.** The validation study showed that the *mBalance* application detects balance losses with a sensitivity of 80% and specificity of 87%. The app's usability was rated favorably. The findings indicate that clear and concise instructions on phone placement and which position to take in front of the cameras are crucial for a successful assessment. **Conclusion.** This publication shows that the presented application is a promising approach to at-home balance assessments using commodity devices. It detects balance losses with high accuracy. Its good usability allows ordinary persons to perform balance assessments at home.

Contributions

The author of this dissertation is the lead/main author of this publication and contributed to it in various aspects such as the idea and conceptualization of the project, writing and proof-reading of the paper, and creation of the visualizations. The publication is based on the bachelor's thesis of Ms. Céline Madeleine Aldenhoven. The thesis project was initiated, conceptualized and supervised by the author of this dissertation. The author of this dissertation supported during the implementation of the *mBalance* application, laboratory experiments, and data analysis. She contributed significantly to the writing, creation of the visualizations used in this publication, and proof-reading.

dHealth 2022
G. Schreier et al. (Eds.)

doi:10.3233/SHTI220344

mBalance: Detect Postural Imbalance with Mobile Devices

Céline Madeleine ALDENHOVEN[a,1], Lara Marie REIMER[a,b,2] and Stephan JONAS[b]
[a] *Technical University of Munich, Garching, Germany*
[b] *University Hospital Bonn, Bonn, Germany*

Abstract. Background: Postural imbalance can be adopted for the early detection of age-related diseases or monitoring the course of the disease treatment; especially in monitoring, frequent balance measurement is crucial. This is mainly done through regular in-person examinations by a physician currently. Feedback in between examinations is often missing. Objectives: This paper proposes mBalance, a mobile application that uses the Romberg test to detect postural imbalance. mBalance provides a camera-based, low-cost approach to measure imbalance frequently at home using mobile devices. Methods: Imbalance detection accuracy and usability was evaluated in two separate studies with 31 and 30 participants, respectively. Results: mBalance correctly detected imbalance with a sensitivity of 80% and a specificity of 87%. The study found good usability with no significant problems. Conclusion: Overall, this study solves the problem of postural imbalance detection by digitizing a validated balance test into an easy-to-use mobile application.

Keywords. mHealth, Mobile Applications, Augmented Reality, Telemedicine, Postural Balance

1. Introduction

Postural imbalance occurs in several diseases and can be an early indicator. It is also useful for progress monitoring, severity assessment, and prognosis. According to Hugues et al., "**Stroke** frequently results in balance disorders" [1]. Balance has also been studied with **Parkinson's disease** [2, 3, 4]. Weiss et al. suggest using the balance test results as an additional and objective measure for tracking the progression of the disease and for supporting its treatment [5]. Melillo et al. pointed out the importance of balance detection in a non-clinical environment for **Multiple Sclerosis** patients to identify the severity of postural imbalance during a clinical checkup [6]. **Cognitive impairment, brain tumors,** or **injuries** can also lead to postural imbalance [1, 7, 8, 9, 10].

Furthermore, postural imbalance is often linked to an increased risk of falling, which is a common cause of death and disability in older adults [11]. Tracking postural imbalance progress in older adults can help doctors identify interventions to help increase the patient's balance most effectively and thus, reduce falls risk. Identifying balance high-risk falls behaviors is crucial for older adults' health and to destress the health care system. Additionally, assessment methods are necessary to implement preventative measures effectively [12]. Heitmann et al. found that women with a history of falls

[1] Corresponding Author: Céline Madeleine Aldenhoven, E-Mail: celine.aldenhoven@tum.de
[2] Corresponding Author: Lara Marie Reimer, E-Mail: reimer@tum.de

performed significantly worse in the balance tests than non-fallers [13]. Roeing shows that mobile balance detection is promising as smartphones can distinguish between high- and low-risk older adult fallers [14].

Postural imbalance assessment is usually performed during a medical examination to give the physician a snapshot of the patient's current condition. Other allied medical professionals, such as physiotherapists, occupational therapists, and sports scientists, administer balance assessments as well. The Romberg test is commonly used in the clinical field to measure static balance [15, 16, 17]. This test measures the subject's balance while standing straight with eyes opened and closed. Response is scored by noting the amount of time balance is maintained in this posture. Patients must track their balance at home for balance tests that require extended periods, which is often subjective. An easy-to-use mobile application that enables patients to conduct accurate balance tests at home changes this.

Even though postural imbalance detection is essential in early disease detection and progress supervision, no mobile application exists yet that allows patients to detect postural imbalance early or supplies doctors with balance data to track the progress of their treatment. Modern smartphones are equipped with various sensors, including optical and inertial. This enables new, accessible and low-cost clinical balance measurements [18]. Research shows that detecting postural imbalance through sensors integrated into mobile devices, i.e., accelerometers, is feasible. Galan-Merchant et al. used the accelerometer of an iPhone 4 to analyze the kinematics of the Romberg test between frail and non-frail older adults [19]. Fleury et al. developed a system included in a smartphone to help the user keep balance using audio-biofeedback [20]. Galan-Merchant et al. stated that the integrated inertial sensors in an iPhone 4 are sufficiently reliable and accurate to evaluate and identify kinematic patterns in balance tests [21]. Ozinga et al. used iPads to show that mobile devices provide an accurate and reliable data output to quantify the balance of Parkinson's disease patients [22, 23]. Overall, the accelerometer is the most used sensor in literature that assesses balance and human motion [5, 24 – 28].

While using the accelerometer of a mobile phone is convenient for conducting the Romberg test, it cannot provide any insights into whether the test was performed correctly. This is because there is no indication of whether a person was standing in the correct position or even standing at all. A more detailed balance analysis is possible through video-based motion tracking [29]. For example, Romaniszyn et al. [27] studied a marker-less, video-based balance state estimation of older adults. Another example is a low-cost, video-based tool performing clinical gait analysis using a marker-based gait assessment with five markers on the most important landmarks of the leg [28].

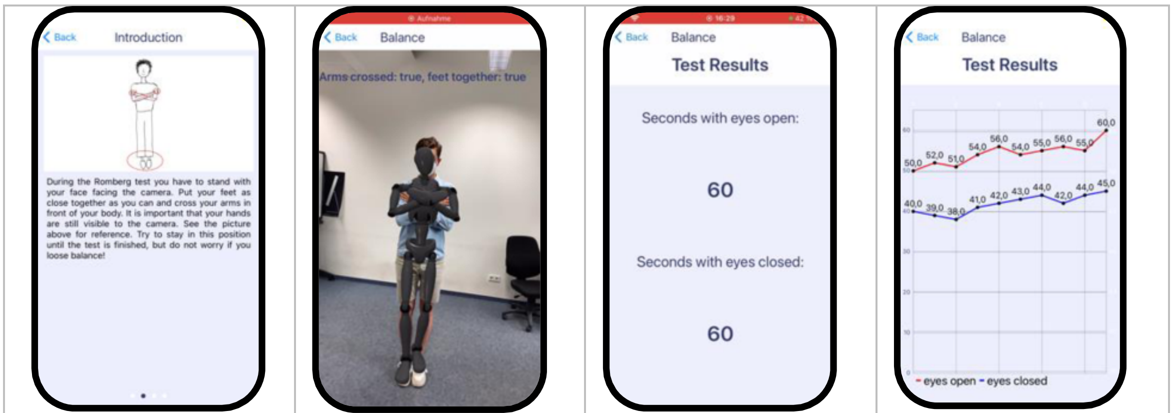
Our research is an essential step in moving clinical tests to mobile devices. We propose *mBalance*, a video-based mobile application that uses augmented reality body tracking to perform and evaluate the Romberg test. *mBalance* enables users to perform a balance assessment at home, and can be used to provide frequent feedback to their supervising physicians. The application guides the user through the assessment and validates the correct execution through video recording.

We performed two studies. In **validation study I)**, we evaluated the detection accuracy of the augmented reality body tracking approach. The **usability study II)** analyzed the app's usability, following the thinking aloud method by Nielsen [30].

2. Methods

2.1. The mBalance Application

The goal of *mBalance* is to provide an easy-to-use mobile application to perform the Romberg test at home. The app uses an iPhone or iPad and the ARKit framework for computer-vision-based pose detection of the human body. First, a user is guided through the test instructions (Figure 1a). The app will then automatically detect the patient's position (Figure 1b). Loss of balance, meaning leaving the test position, can also be detected. Furthermore, the app allows progress monitoring (Figure 1c-d).



a) Introduction Page: provides information on how to perform the test.

b) Test view of mBalance: shows the tracked body as a skeleton overlay in the correct test position.

c) Test results view: shows the results of the finished test.

d) Test results of all previously recorded tests: shows the inclination of the test results as a graph.

Figure 1. Main screens of the *mBalance* application.

mBalance conducts the Romberg test according to the specifications by [16, 19], with the arms crossed in front of the chest and feet together. The test is performed twice, once with open eyes and once with closed eyes. Each test run takes a maximum of 60 seconds. After the first test run, the patient is instructed to close their eyes and repeat the test via audio commands. If the patient loses balance in one of the test runs, time is stopped. Assessment continues with the second run or completion of the balance test. The test results are given in number of seconds for open and closed eyes with a result of 60 seconds indicating no balance impairments.

In *mBalance*, the detection algorithm recognizes whether the subject is standing in the correct test position or losing their balance by calculating the difference between the hand and feet coordinates. The origin of ARKit's body coordinate system is the pelvis center. The x-Axis is aligned through the pelvis with positive coordinate values toward the right side of the pelvis, and negative coordinates toward the left side. The y-Axis is aligned toward the head, with positive values toward the direction of the head and negative coordinates toward the feet in a standing position. The z-Axis is perpendicular to the pelvis and the vector between the pelvis and head. The coordinate system changes its orientation based on the movement of these landmarks. The *mBalance* algorithm calculates the correct hand position based on the distance between hands and feet based on this coordinate system. It additionally verifies that the hands are in a crossed position, meaning that the left hand is on the right side of the root coordinate and the right hand is on the left side.

The correct detection of the body and the extremities of ARKit can be affected by several reasons: Bad lighting conditions, hidden parts of the body, or the camera position, which is a known issue for computer vision-based motion capture systems [31]. We filter out poses where the algorithm detected position change for less than *q* seconds, assuming that this was a wrongly detected movement, to improve accuracy as this could otherwise lead to an erroneous balance detection result. The parameter *q* depends on the device's performance, as ARKit benefits from higher processing power and should be adapted to the device used. The value *q* has to be adjusted just right, as a higher *q* can result in a noticeable delay of balance loss detection, while a smaller *q* could incorrectly detect a loss of balance.

In the pre-tests of the study, we empirically determined the thresholds for verifying the test position detection and the parameter *q* by calculating the mean values for a correct detection based on the test data. The code was adapted accordingly to run on an iPad Pro (11" version, 2021, Apple Inc, Cupertino, USA). Our application's code can be found on GitHub[3] to support future research.

2.2. Participants, Study Setting, and Data Processing

The validation study **I)** and usability study **II) included** 31 participants (5 female) and 29 participants (4 female) without balance impairments, respectively. The participants were between 19 and 56 years old. The balance of the subjects was tested before the study. Due to technical difficulties, one dataset was excluded in each study, leading to 30 and 28 usable datasets for the two studies, respectively.

In validation study **I)**, each subject conducted the balance test three times; two times while keeping their balance and one time while intentionally losing balance. The subjects could randomly choose when and how often to lose their balance by taking a step, opening their arms, or both. Each subject was instructed before the study started on the correct test position, with feet together and hands crossed and visible to the camera. The tests were conducted on an iPad Pro 11 (M1 chip, LiDAR sensor). The iPad was placed on a tripod and kept at the same distance and height for all subjects. The participants stood close to a white wall, and wore tight clothes to increase body detection accuracy. The study conductor assessed whether the test results for each participant was detected correctly by *mBalance*. A true loss of balance was determined by the study conductor if the participants left the test position through opening their arms or taking a step; otherwise, no loss of balance was indicated. We quantitatively analyzed the results.

For study **II)**, the participants were interviewed on-site using the thinking aloud technique [30]. The subjects were told: "*This app conducts a balance test. Your goal is to do the test once.*" No further information was provided. The subjects used an iPhone SE (2nd Generation, Apple Inc, Cupertino, USA) and a tripod. We grouped similar answers into categories (Table 1) and qualitatively analyzed the feedback.

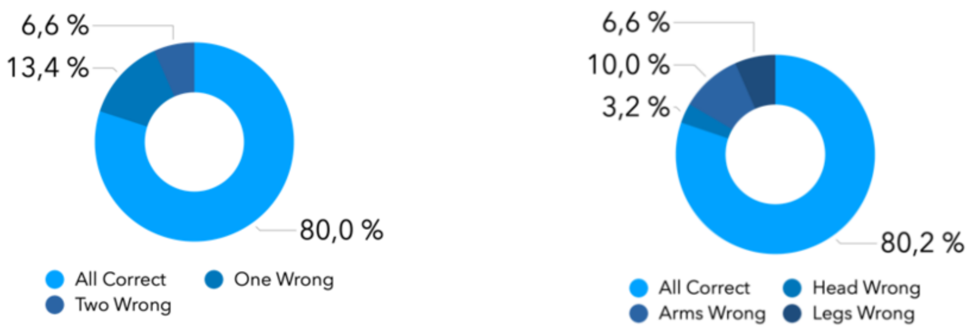
[3] https://github.com/dhg-applab/mBalance

Table 1. The feedback categories for the usability study, grouped into feedback concerning the test setup ((1)–(5)) and the content of the *mBalance* app ((6)–(8)).

Category	Definition
(1) Correct test position	The subject has correctly placed themselves with arms crossed, hands still visible, and feet together in front of the camera.
(2) Correct camera position	The subject correctly used the rear-facing camera.
(3) Placed phone after introduction	The subject read all the introduction pages before placing their phone on the tripod.
(4) Distance to camera clear	The subject did not ask any questions about the distance that they should have to the camera.
(5) Phone position clear	The subject did not have questions about placing the phone horizontally or vertically.
(6) Understood swiping	The subject correctly understood the swiping process in the introduction pages.
(7) Intro test is understandable	The subject had no complaints about the test on the intro pages.
(8) Pictures are helpful	The subject did not have any recommendations about the pictures displayed on the introduction pages.

3. Results

In validation study **I)**, *mBalance* detected all three consecutive measurements correctly for 66% of all participants. Detailed results are presented in Figure 2.



a) The percentages of whether the subsequent balance tests of one participant have been detected correctly.

b) The percentages of whether all losing balance tests have been detected correctly and, if not, which body part has not been detected correctly.

Figure 2. Different metrics regarding the detection rates of the balance tests using *mBalance*.

The contingency table indicates good performance (Table 2). Newer technology increases the precision of body detection, as body detection works better with the iPad 11 Pro than on the iPhone SE. The biggest reason for failed tests was false detection of the hands or feet once detection failed because the subject turned their head. This led to a rotation of the body model, which failed to correctly detect the hands and feet position. This is a problem contained in the ARKit framework.

Finding 1: *mBalance* can measure balance following the Romberg test with good precision. Limitations of the underlying technology caused most errors.

Table 2. Contingency table of the validation test of the *mBalance* application.

		Balance Test Outcome		
		Balance Loss	*No Balance Loss*	*Total*
mBalance Result	*Lost Balance Detected*	24	8	32
	Held Balance Detected	6	52	58
	Total	30	60	90

*Sensitivity = 24 / 30 = 0.8, specificity = 52 / 60 = 0.87.

Results from the usability study **II)** (Table 3) indicated that subjects that did not position themselves correctly (1) either by not putting their feet together perfectly or crossing their arms but did not show their hands. As a high percentage of people positioned themselves correctly, this indicates the suitability of the introduction page. The most important finding in this usability study was the distance that the participants should have the camera (4) was unclear to 37.9% of the subjects.

Finding 2: Providing users with clear and concise instructions on how to position their phones and themselves is crucial for the test's success.

Participants who did not understand the swiping process immediately (6) commented that a “next” button would be beneficial. The subjects (17.2%) were bothered by the amount of text on the introduction pages (7). Most of them suggested using bullet points or animated videos. The pictures that have been shown in the introduction (8) were considered helpful by 93.1%. One thing which could improve the pictures was the ability to zoom in the images for a better view.

Finding 3: Using short and well-ordered descriptions and visualizations should be preferred over detailed textual descriptions as an introduction for the user.

Table 3. Results of the usability study for the different feedback categories are presented in Table 2.

Category	% of Participants	Category	% of Participants
(1) Correct test position	75,9%	*(5) Phone position clear*	90%
(2) Correct camera position	82,8%	*(6) Understood swiping*	89,7%
(3) Placed phone after introduction	69%	*(7) Intro text is understandable*	82,8%
(4) Distance to camera clear	62,1%	*(8) Pictures are helpful*	93,1%

4. Discussion and Future Work

4.1. Limitations

Both studies **I)** and **II)** had a small study group of 31 and 29 participants, respectively, leading to non-representative statistics. This limited group size was necessary due to COVID-19 restrictions. Additionally, *mBalance* is a limited application. There is room for complexity, like different variations of the Romberg test (sharpened, single leg), their specific consequences, and support for diagnosis.

In study **I)**, everyone who participated was healthy. Therefore, the detection of balance loss had to be deliberately simulated, which differs from patients suffering from postural imbalance. Another study should be conducted on subjects with neurological disorders to assess *mBalance* with the target group. Our study was conducted under ideal circumstances (good lighting, white wall on the background). In less than ideal conditions, human body detection might be challenging, which affects the accuracy of the app. In our validation study, we instructed all subjects on how to position themselves correctly. In the app's actual use case, more human subjects' errors need to be considered. Additionally, no security measures were implemented owing to the participants healthy balance state.

In study **II)**, it is possible that subjects did not speak all their thoughts out aloud, even though this was encouraged. This would imply that the results presented in this research might deviate from reality.

4.2. Imbalance Detection Accuracy

To overcome current inaccuracies in Apple ARKit's body tracking, we selected a time interval q that determines whether a movement is caused by a body detection inaccuracy or an actual movement of the recorded person. This time interval was selected to fit the device's computing capabilities and its errors in detecting the human body to ensure the best performance of the application. In future work, this could be improved to ensure transferability to other devices, for example, by applying filters to the signals emitted by ARKit.

In our study, *mBalance* achieved good results for state-of-the-art research in digital health. This implies that we can enable patients to test themselves at home, even though mobile body detection is still error prone. This would allow patients to perform regular assessments easily and monitor their balance, leading to early detection of any change.

4.3. Usability Discussion

Most of the participants positioned themselves correctly, but this could be improved by highlighting the explanation of the test position during the introduction. Another way to ensure the correct position is to enhance the pose detection algorithm. The accuracy of the pose detection is currently highly influenced by parameters like background and lighting, which require filtering of the motion data. While the performance of the ARKit framework presently limits this, it could be an option for future research projects with other technological approaches or newer technology.

It is essential to tell the user when they are recognized to indicate whether they positioned themselves at a sufficient distance to the camera. However, ARKit still recognizes people even when they are not entirely in the camera frame by calculating the missing coordinates, which might influence the correct balance test result. This could be addressed by providing them with a concrete number of steps they need to walk away from the camera to perform the test during the introduction or automatically calculate whether the calculated coordinates are within the visible coordinate system of the camera. Surprisingly, over 60% of the subjects had no problem with this issue. However, ensuring an adequate distance between the camera and the tracked subject should be addressed in further research.

4.4. Future Work

Our research focused on evaluating the feasibility of balance detection using optical motion tracking methods like Apple ARKit.

The Romberg test includes further variations which should be considered in future implementations of the mobile Romberg test. These include asking the subject to raise their arms in front of them and to gently “push” the subject in different directions, to assess how their balance reacts to external disturbances. Other possible extensions to *mBalance* are balancing in step stance, tightrope stance, or single leg stance. *mBalance* could be further extended by dynamic balance scales like the Berg Balance Scale, which is a commonly applied balance test battery in rehabilitation and falls prevention.

mBalance is currently intended for patients with mild balance impairments as no security measures were implemented. Future work should focus on testing its accuracy with safety measures like having supporting persons next to them to prevent falls in more severely impaired individuals.

5. Conclusion

mBalance suggests using augmented reality body tracking for in-home postural imbalance detection using commodity hardware like smartphones and tablets. The *mBalance* app showed good usability and accuracy of measurement. Even though the underlying technology exhibits inaccuracies in tracking the human body through a smartphone's camera, the approach enables new possibilities of remotely supervising patients and thus, gaining better insights into their health status between appointments. Future work should focus on improving the accuracy of the imbalance detection, and evaluate the approach in a clinical setting.

References

[1] A. Hugues, J. Di Marco, P. Janiaud, Y. Xue, J. Pires, H. Khademi, M. Cucherat, I. Bonan, F. Gueyffier, G. Rode, Efficiency of physical therapy on postural imbalance after stroke: study protocol for a systematic review and meta-analysis, *BMJ open* 7, 1 (2017), e013348.

[2] J.W. Blaszczyk, R. Orawiec, D. Duda-Klodowska, G. Opala, Assessment of postural instability in patients with Parkinson's disease, *Exp Brain Res* 183, 1 (2007), 107–114.

[3] A. Salarian, F.B. Horak, C. Zampieri, P. Carlson-Kuhta, J.G. Nutt, K. Aminian, iTUG, a sensitive and reliable measure of mobility, *IEEE T Neur Sys Reh* 18, 3 (2010), 303–310.

[4] A. Salarian, C. Zampieri, F.B. Horak, P. Carlson-Kuhta, J.G. Nutt, K. Aminian, Analyzing 180° turns using an inertial system reveals early signs of progression of Parkinson's disease, *IEEE Eng Med Bio* (2009), 224–227.

[5] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, I. Maidan, N. Giladi, T. Gurevich, J.M. Hausdorff, Can an accelerometer enhance the utility of the Timed Up & Go test when evaluating patients with Parkinson's disease?, *Med Eng Phys* 32, 2 (2010), 119–125.

[6] F. Melillo, A. Di Sapio, S. Martire, M. Malentacchi, M. Matta, A. Bertolotto, Computerized posturography is more sensitive than clinical Romberg Test in detecting postural control impairment in minimally impaired Multiple Sclerosis patients, *Mult Scler Relat Dis* 14 (2017), 51–55.

[7] A.L. Betker, T. Szturm, Z.K. Moussavi, C. Nett, Video game-based exercises for balance rehabilitation: a single-subject design, *Arch Phys Med Rehab* 87, 8 (2006), 1141–1149.

[8] J.-A. Gil-Gómez, R. Lloréns, M. Alcañiz, C. Colomer, Effectiveness of a Wii balance board-based system (eBaViR) for balance rehabilitation: a pilot randomized clinical trial in patients with acquired brain injury, *J Neuroeng Rehabil* 8, 1 (2011), 1–10.

[9] R. Llorens, C. Colomer-Font, M. Alcañiz, E. Noé-Sebastián, BioTrak virtual reality system: effectiveness and satisfaction analysis for balance rehabilitation in patients with brain injury, *Neurologia (English Edition)* 28, 5 (2013), 268–275.
[10] A. Mirelman, A. Weiss, A.S. Buchman, D.A. Bennett, N. Giladi, J.M. Hausdorff, Association between performance on Timed Up and Go subtasks and mild cognitive impairment: further insights into the links between cognitive and motor function, *J Am Geriatr Soc* 62, 4 (2014), 673–678.
[11] B.R. Greene, E.P. Doheny, R.A. Kenny, B. Caulfield, Classification of frailty and falls history using a combination of sensor-based mobility assessments, *Physiol Meas* 35, 10 (2014), 2053.
[12] K.L. Roeing, K.L. Hsieh, J.J. Sosnoff, A systematic review of balance and fall risk assessments with mobile phone technology, *Arch Gerontol Geriat* 73 (2017), 222–226.
[13] D.K. Heitmann, M.R. Gossmann, S.A. Shaddeau, J.R. Jackson, Balance performance and step witdth in noninstitutionalized, elderly, female fallers and nonfallers, *Phys Ther* 69, 11 (1989) 923-931.
[14] K.L. Roeing, The assessment of standing balance with smartphone technology for the prevention of falls in older adults (2017).
[15] R.C. Briggs, M.R. Gossman, R. Birch, J.E. Drews, S.A. Shaddeau, Balance performance among noninstitutionalized elderly women, *Phys Ther* 69, 9 (1989), 748–756.
[16] S.K. Hong, J.H. Park, S.Y. Kwon, J.-S. Kim, J.-W. Koo, Clinical efficiacy of the Romberg test using a foam pad to identify balance problems: a comparative study with the sensory organization test, *Eur Arch Oto-Rhino-L* 272, 10 (2015), 2741-2747.
[17] N.S. Longridge, A.I. Mallinson, Clinical Romberg testing does not detect vestibular disease, *Otol Neurotol* 31, 5 (2010), 803–806.
[18] N.M. Kosse, S. Caljouw, D. Vervoort, N. Vuillerme, C.J.C. Lamoth, Validity and reliability of gait and postural control analysis using the tri-axial accelerometer of the iPod touch, *Ann Biomed Eng* 43, 8 (2015), 1935–1946.
[19] A. Galán-Mercant, A.I. Cuesta-Vargas, Mobile Romberg test assessment (mRomberg), *BMC research notes* 7, 1 (2014), 1–8.
[20] A. Fleury, Q. Mourcou, C. Franco, B. Diot, J. Demongeot, N. Vuillerme, Evaluation of a Smartphone-based audio-biofeedback system for improving balance in older adults-A pilot study, *IEEE Eng Med Bio* (2013), 1198–1201.
[21] A. Galán-Mercant, F.J. Barón-López, M.T. Labajos-Manzanares, A.I. Cuesta-Vargas, Reliability and criterion-related validity with a smartphone used in timed-up-and-go test, *Biomed Eng Online* 13, 1 (2014), 1–11.
[22] S.J. Ozinga, S.M. Linder, J.L. Alberts, Use of mobile device accelerometry to enhance evaluation of postural instability in Parkinson disease, *Arch Phys Med Rehab* 98, 4 (2017), 649–658.
[23] S.J. Ozinga, A.G. Machado, M.M. Koop, A.B. Rosenfeldt, J.L. Alberts, Objective assessment of postural stability in Parkinson's disease using mobile technology, *Movement Disord* 30, 9 (2015), 1214–1221.
[24] S. Mellone, C. Tacconi, L. Chiari, Validity of a Smartphone-based instrumented Timed Up and Go, *Gait Posture* 36, 1 (2012), 163–165.
[25] M. O'Sullivan, C. Blake, C. Cunningham, G. Boyle, C. Finucane, Correlation of accelerometry with clinical balance tests in older fallers and non-fallers, *Age ageing* 38, 3 (2009), 308–313.
[26] L. Palmerini, S. Mellone, G. Avanzolini, F. Valzania, L. Chiari, Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test, *IEEE T Neur Sys Reh* 21, 4 (2013), 664–673.
[27] P. Romaniszyn, J. Kawa, P. Stępien, A. Nawrat-Szołtysik, Video-based time assessment in 360 degrees turn Berg balance test, *Comput Med Imag Grap* 80 (2020), 101689.
[28] P. Soda, A. Carta, D. Formica, E. Guglielmelli, A low-cost video-based tool for clinical gait analysis, *IEEE Eng Med Bio* (2009), 3979-3982.
[29] L.M. Reimer, S. Weigel, F. Ehrenstorfer, M. Adikari, W. Birkle, S. Jonas, Mobile Motion Tracking for Diseases Prevention and Rehabilitation Using Apple ARKit. *St Heal T* (2021).
[30] J. Nielsen, *Usability engineering*, Morgan Kaufmann, 1994.
[31] M. Zago, M. Luzzago, T. Marangoni, M.D. Cecco, M. Tarabini, M. Galli, 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology*, *8,* (2020).

7 Publication D: Developing an App for Cardiovascular Prevention and Scientific Data Collection[1]

This publication presents the concept of an application to prevent Cardiovascular Diseases (CVDs) by promoting physical activity to lower the individual risk of CVDs among other lifestyle changes. Moreover, the application's architecture allows identifying and recruiting user cohorts for scientific studies. It enables the secure collection of quantitative activity data from the smartphone and connected mobile devices. Since publication of the app's concept in May 2021, the app was developed, evaluated, and released to the public in May 2022 [135, 136].

Conference: dHealth 2021 - Navigating healthcare through challenging times
Number of Pages: 2
Type: Short Paper
Review: Peer Reviewed (3 Reviewers)

Summary

Problem. CVDs are the leading cause of death worldwide. Many risk factors for such diseases exist, including unhealthy habits such as smoking, an unhealthy diet, or lack of physical activity. While many apps exist that target individual risk factors, little data is available on the utilization, perception, and long-term effects of such apps on CVD prevention. **Objective.** In this publication, we present the concept of a mobile app that aims at being a digital life coach for the prevention of CVD. The app's goal is to reduce the user's risk of CVD and perform scientific studies while preserving users' data privacy. **Methods.** To validate the concept of the app, a prototype of the app will be built. This prototype will be used for usability studies on how the target group perceives such an application. **Results.** This research expects to determine whether the

[1]The following publication is embedded for additional context and has not been considered as part of the assessment of this dissertation.

targeted user group would be interested in using such an application and thus reach a broad enough user base to perform scientific data collection. In addition, we expect to develop a concept to collect scientific data and protect it sufficiently. **Conclusion.** We propose a mobile application to promote CVD prevention. The mobile application should serve as a life coach and as a platform for scientific data collection at the same time. We expect to achieve high acceptance rates of the concept in validation studies.

Contributions

The author of this dissertation is the lead/main author of this publication. As such, she contributed substantially to the writing of this publication, including the literature review and design of the concept described in the publication. Beyond the publication, the author of this dissertation managed the technical implementation of the concept and the distribution to the App Stores.

Navigating Healthcare Through Challenging Times
D. Hayn et al. (Eds.)

doi:10.3233/SHTI210095

Developing an App for Cardiovascular Prevention and Scientific Data Collection

Lara Marie REIMER[a,1], Fabian STARNECKER[a,b,c], Heribert SCHUNKERT[a,b,c] and Stephan JONAS[a]
[a] *Technical University of Munich, Munich, Germany*
[b] *German Heart Centre Munich, Munich, Germany*
[c] *German Centre for Cardiovascular Research, Berlin, Germany*

Abstract. Background: Mobile apps may encourage a lifestyle that avoids unhealthy behaviors, such as smoking or poor nutrition, which promotes cardiovascular diseases (CVD). Yet, little data is available on the utilization, perception, and long-term effects of such apps to prevent CVD. Objectives: To develop a mobile app concept to reduce the individual CVD risk and collect information addressing research questions on CVD prevention while preserving data privacy and security. Methods: To validate the concept, a prototype will be built, and usability studies will be performed. Results: We expect to determine whether it is possible to reach a broad user base and to collect scientific information while protecting user data sufficiently. Conclusion: To address CVD prevention, we propose a mobile coaching app. We expect high acceptance rates in validation studies.

Keywords. mHealth, Mobile Applications, Cardiovascular Diseases

1. Introduction

Cardiovascular diseases (CVD) remain the leading cause of death in the Western world [1]. Apart from conventional risk factors, such as smoking or hypertension, unhealthy lifestyle, including poor nutrition and psychological stress, are known to increase CVD risk. Modern technologies inform about risk factors and coach their users to adhere to a healthier lifestyle. Studies have shown that mobile medical apps can positively impact daily behavior in the short term [2,3]. Lifestyle apps target a broader user base while focusing on a subset of risk factors [4]. To evaluate the impact of both application types on CVD risk, long-term research data is necessary. Yet, such data collections are limited by the high costs of continuous data acquisition. Moreover, there is a need to evaluate the effects of mobile apps in primary prevention. Previous trials mainly included subjects with diagnosed CVD or other high-risk related diseases [2,3].

2. Methods

This paper presents a mobile app concept, which aims to provide a lifestyle application that targets relevant risk factors for CVD. Moreover, the concept allows collecting

[1] Corresponding Author: Lara Marie Reimer, Technical University of Munich, Germany, E-Mail: reimer@tum.de

scientific data through the app on how lifestyle changes impact CVD risk. We will build a prototype to validate the concept in usability studies.

3. Results

The system architecture consists of multiple subsystems. Health information and study questionnaires are distributed via a content delivery network to a web platform and a mobile app. The web platform serves as an entry point into the app. The app provides personalized content for the user and gathers individual health information. All data is processed locally on the device, ensuring maximum data security and protection. Health data, collected via smartphone sensors and connected wearables, is automatically imported into the app. Additional information can be entered manually. Both data types include parameters used for the calculation of the individual risk for CVD.
The mobile app decides based on prerequisites whether a user is eligible for a study. Studies can specify additional data to be collected. Only upon user consent, the app transmits anonymized health data to the study database. The data transmission is secured using end-to-end encryption between the study database and the user's device. Each study is examined for suitability of conduction in the app and requires an ethical vote.

4. Discussion and Future Work

To address CVD prevention, we propose a mobile app that allows the app user to lower their individual risk. For gaining insights into prevention, the app offers an anonymous collection of clinical data from users. The development of the prototype will be finished in 2021. Usability studies will be performed, emphasizing whether we can establish a user base broad enough for trials and whether the app protects user data sufficiently. Due to the focus on data protection, we expect high acceptance rates of the app.

Funding

Funded by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digmed-bayern.de)

References

[1] P. Bhatnagar, K. Wickramasinghe, E. Wilkins, N. Townsend, Trends in the epidemiology of cardiovascular disease in the UK. *Heart*, **102**(24) (2016), 1945-1952.
[2] G.M. Coorey, L. Neubeck, J. Mulley, J. Redfern, Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. *Eur J Prev Cardiol* **25**(5) (2018), 505-521.
[3] L. Pfaeffli Dale., R. Dobson., R. Whittaker, R. Maddison, The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. *Eur J Prev Cardiol* **23**(8) (2016), 801-817.
[4] G. Flores Mateo, E. Granado-Font, C. Ferré-Grau, X. Montaña-Carreras, Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. *J Med Internet Res* **17**(11) (2015), e253.

8 Summary

This dissertation provided an overview of the possibilities and limitations of mobile motion capture technologies in mobile health and fitness applications, particularly optical 3D motion capture. However, several open topics remain. In this concluding section, we discuss several open topics, give an outlook on future work, and summarize the conclusions of this research.

8.1 Discussion

The field of physical activity recognition using mobile devices is wide. In this research, we focused on exploring optical, mobile 3D motion capture frameworks on the example of Apple *ARKit*. Essential thoughts on the topic are reflected in the following discussion. This includes the choice of *ARKit* as an exemplary framework over other available 3D motion capture frameworks, which would be available on a broader range of devices. Camera-based motion capture raises data privacy concerns, as users have to film themselves. This concern is aggravated as the filming happens in a potentially private setting. It also needs to be considered that, while many people enjoy the benefits of controlling their health data, others prefer to keep tracking their health at a minimum. Users will likely only use health and fitness applications for a more extended time if they present a clear motivation. The discussion is completed by identifying possible threats to validity concerning the research design.

8.1.1 Availability of Mobile Motion Capture Technologies

In this dissertation, we selected the Apple *ARKit* framework as an example for 3-dimensional motion capture methods, allowing local real-time optical sensor data processing. *ARKit* is a promising framework easily accessible to developers and available on many devices. It is well-integrated into the iOS and iPadOS ecosystems, facilitating the support of different sensor systems on these platforms, such as Light Detection And Ranging (LiDAR). However, it is only available on the iOS and iPadOS platforms, limiting its availability to 28% of smartphone users worldwide [137]. It is, therefore, reasonable to consider alternative approaches for mobile motion capture, which would be available to both iOS/iPad OS and Android users. With the LiDAR technology, Apple integrated a powerful depth sensor in their smartphones, which is used to enhance object detection in the *ARKit* framework. The most common 3D motion capture framework available on Android, *MediaPipe Pose*, provides similar functionality as Apple *ARKit* while being available on iOS, Android, and even web applications. The portability comes at the cost that *MediaPipe Pose* does not support additional sensors available on a small subset of devices, so it only uses the video data. While some Android devices are equipped with depth sensors, they are not used by the most common motion capture frameworks available on Android. Therefore, potential alternatives for optical motion capture include mobile motion capture frameworks that only calculate depth information through the RGB image. Furthermore, little research is available on the accuracy of *MediaPipe Pose* for mobile 3D motion capture.

A prerequisite for optical motion capture algorithms is the sufficient processing power of the computer-vision computations device. With their M1 processor, Apple released a powerful processor integrated into newer devices [134], which further motivated the use of *ARKit* in this dissertation.

8.1.2 Data Privacy in Camera-based Motion Capture

ARKit and similar technologies mainly rely on Red-Green-Blue (RGB) camera pictures for motion capture. While that resembles an easily accessible method, it requires the users to film themselves. The motion capture can happen in private environments such as the users' homes and might require them to wear tight or short clothing. The captured motion could also consist of movements that the users would prefer not to be observed while doing them. All of the mentioned problems underline the importance of securely handling the user data and ensuring they are sufficiently protected.

We specifically selected *ARKit* as a technology that can process the recorded RGB images on the device without the necessity to send them to a server for extracting the joint information. This enables secure processing of the raw image data, increasing the barriers to mapping the data onto a specific person. Even though movement patterns can be classified as biometric data [138], the raw camera pictures and videos do not need to be stored if later processing or synchronization of the data across multiple devices is desired. However, specific settings prevent the complete anonymization of personal data, such as conducting scientific studies or environments in which users want to share their data with a third party or synchronize their data across multiple devices. In these cases, a secure infrastructure must be established to protect sensitive user data. In CS 5, we presented an approach to performing scientific studies and securely collecting and storing sensitive user data. However, the current possibilities of protecting user data still expose vulnerabilities. As in every software, it should therefore carefully be considered which data needs to be stored on which part of the system.

8.1.3 User Motivation and Perception of Health Tracking

This dissertation showed the technical potential of physical activity recognition and analysis. However, the full potential of motion capture applications is not only defined by their technical features but also the application's usability and the users' long-term motivation. The full potential depends on how the target users perceive such apps and their motivation to use them. Physical activity recognition in sports training offers a clear motivation: regular use can improve the trainee's performance and prevent injuries caused by wrong techniques. Health applications often do not offer clear motivation as their focus is on long-term supervision and lifestyle changes. Therefore, embedding health applications into a context that delivers long-term motivation is essential. Such additional sources of motivation can include communication with supervising physicians or bonus programs offered by several insurance companies. Another commonly used aspect is gamification, which could be integrated into such applications to facilitate motivation and add entertaining content.

Besides the motivation, the user's perception of tracking their health data is equally important. In this context, the user's perception refers to their interest in tracking health data and getting to know more about their physical and mental health. While many smartphone users enjoy tracking their health parameters, others might object to gathering these amounts of health data and prefer not to track them. Such users are hard to convince to use mobile applications to track their health data.

8.1.4 Limitations

Several factors impact the **external validity** of this research. The *population validity* is influenced due to the focus on technical feasibility and usability, which was evaluated on healthy participants. To counteract population validity, we aimed to include participants from the respective target populations in several parts of the research. We included healthy amateur athletes in the validations of the *ARKit* framework in Research Question (RQ) 2 and performed the Case Studys (CSs) 1 and 3 with representatives of goalkeepers and golf trainers. However, in CSs 4 and 5, we did not include participants diagnosed with postural imbalance or Cardiovascular Disease (CVD) due to the focus on a technical validation of the systems.

The different experiments and CSs only included a small and relatively homogeneous number of participants, making this research prone to *selection bias*. While we aimed at balancing the number of male and female participants, the experiments and CSs only included subjects between 20 and 30 years with light skin color. To increase the validity of the *ARKit* validation concerning the selection bias, we performed two subsequent experiments with different subjects and exercises. Moreover, we applied the approach to different CSs with up to 30 participants, targeting use cases in health and fitness.

The results of this research have only been validated in real-life conditions to a limited extent, impacting the *ecological validity*. The validations of RQ 2 took place in a controlled setting. However, the CSs used in-field settings, especially the recordings of the goalkeeper exercises and the golf jump combinations, which happened during regular training.

As both the laboratory experiments and the CSs 3 and 4 took place in a controlled environment, the results might be influenced by the *experimenter effect*. The participants knew about the study setting and might have behaved or answered differently from those in a natural environment. To limit these effects, several measures were introduced. The studies included repeated movements by several participants, aiming at natural exercise executions. Usability studies were designed to identify usability problems based on user interactions, e.g., using the Thinking Aloud technique. Conclusions were drawn based on observations in CS 3 and 4. In the case of the questionnaire in CS 3, this implied that users were asked to describe which app user interface elements they would use to perform a specific task. In CS 4, they were observed while interacting with the device.

Several threats affect the **internal validity** of this dissertation. All evaluations took place in a controlled setting, limiting the number of external factors influencing the results. We identified factors that could influence the testing results, such as which participant is being captured or which exercise is performed, and analyzed their impact as part of RQ 2.2. However, additional factors which were not investigated as part of the study could have influenced the results, such as the participants' clothing or other attributes.

8.2 Future Work

This research showed the technical possibilities of physical activity recognition using mobile devices. However, several questions remain open and are subject to future work. The presented CSs were designed as technical prototypes. The evaluations included small cohorts consisting of healthy individuals. Future work should focus on evaluations of the correct target group, thus including trainers, athletes, and patients. Especially evaluations on patients require the development of the presented app prototypes as medical software under the Medical Device Regulation (MDR), which should be targeted in future work. Even though the HerzFit app, described as a concept by Publication [D], was already developed and released to the public, a detailed evaluation is still missing and should also be targeted by future work.

We mainly considered Apple *ARKit* as motion capture technology due to its advantages in integration with hardware and performance. As *ARKit* is only available on iOS and iPadOS devices, other mobile motion capture technologies should be researched to enable similar use cases on the Android operating system.

All described use cases only used motion data for activity analysis. Future work could integrate more health data into the use cases and thus generate additional metrics, such as the motion intensity. Integration with additional health data to measure intensity could be especially relevant in use cases, where it is essential to stay at a specific intensity to reach the desired effect, as in a cardio workout.

We investigated five different CSs where physical activity recognition using mobile devices could be beneficial. However, this is only a small subset of possibilities for such applications. Only a few mobile applications using physical activity recognitions exist in the medical field, especially regarding a motion quality assessment. Possible application fields include orthopedics, neurology, or pediatrics, but also use cases in sports and fitness applications. Future work could develop and evaluate additional technical prototypes to investigate the possibilities of physical activity recognition in these fields, followed by more extensive evaluation studies if they appear promising.

8.3 Conclusion

This dissertation evaluated physical activity recognition methods on mobile devices for applicability in sports and health use cases. The scope covered literature research on methods of recognizing physical activity, laboratory experiments on the accuracy of optical motion capture methods on the example of Apple *ARKit* and influencing factors, and the exploration of potential use cases for mobile motion capture in the field of sports and health.

We showed that mobile optical motion capture frameworks are a possibility for evaluating motion quality. New developments in mobile sensors and processors enable on-device processing but have only been evaluated to a limited extent. We contributed to closing this research gap by providing a validation of the Apple *ARKit* framework for mobile 3D motion capture using optical smartphone sensors. We showed that *ARKit* could track a person's motion with good accuracy but with significant deviations compared to a gold standard system. Good tracking results mainly depend on the visibility of the tracked joints and the motion performed. In five different case studies, we showed that the accuracy of the physical activity recognition using frameworks such as *ARKit* is sufficiently accurate to perform activity classification, exercise counting, and balance assessments. We achieved good usability results and good user acceptance among one of the target groups. We also provided an approach to integrating physical activity recognition into a bigger context and leveraging the data for scientific studies while preserving high data privacy standards.

Acronyms

AREA	Augmented Reality Exercise Analysis.
CNN	Convolutional Neural Network.
CS	Case Study.
CVD	Cardiovascular Disease.
DNN	Deep Neural Network.
DTW	Dynamic Time Warping.
EMS	Electromagnetic Measurement System.
GDPR	General Data Protection Regulation.
GPS	Global Positioning System.
HAR	Human Activity Recognition.
HMC	Human Motion Capture.
IMS	Inertial Measurement System.
IMU	Inertial Measurement Unit.
IoT	Internet of Things.
IPS	Image Processing System.
LiDAR	Light Detection And Ranging.
LSTM	Long-Short-Term-Memory Network.
MDR	Medical Device Regulation.
MecMS	Mechanic Measurement System.
OMS	Optical Motion Capture System.
PAF	Part Affinity Field.
RGB	Red-Green-Blue.
RGB-D	Red-Green-Blue-Depth.
RMSE	Root Mean Squared Error.
RQ	Research Question.

SRCC Spearman Rank Correlation Coefficient.

ULS Ultrasonic Localization System.

wMAE weighted Mean Absolute Error.

List of Figures

List of Tables

Bibliography

[1] W. B. Kannel, “Some health benefits of physical activity: The framingham study”, *Archives of Internal Medicine*, vol. 139, no. 8, p. 857, Aug. 1, 1979. DOI: 10.1001/archinte.1979.03630450011006.

[2] F. J. Penedo and J. R. Dahn, “Exercise and well-being: A review of mental and physical health benefits associated with physical activity:” *Current Opinion in Psychiatry*, vol. 18, no. 2, pp. 189–193, Mar. 2005. DOI: 10.1097/00001504-200503000-00013.

[3] M. Reiner, C. Niermann, D. Jekauc, and A. Woll, “Long-term health benefits of physical activity – a systematic review of longitudinal studies”, *BMC Public Health*, vol. 13, no. 1, p. 813, Dec. 2013. DOI: 10.1186/1471-2458-13-813.

[4] J. Midtgaard, M. T. Baadsgaard, T. Møller, B. Rasmussen, M. Quist, C. Andersen, M. Rørth, and L. Adamsen, “Self-reported physical activity behaviour; exercise motivation and information among danish adult cancer patients undergoing chemotherapy”, *European Journal of Oncology Nursing*, vol. 13, no. 2, pp. 116–121, Apr. 2009. DOI: 10.1016/j.ejon.2009.01.006.

[5] C. Visovsky and C. Dvorak, “Exercise and cancer recovery”, *Online Journal of Issues in Nursing*, vol. 10, no. 2, p. 7, Mar. 28, 2005.

[6] L. E. Stephenson, D. G. Bebb, R. A. Reimer, and S. N. Culos-Reed, “Physical activity and diet behaviour in colorectal cancer patients receiving chemotherapy: Associations with quality of life”, *BMC Gastroenterology*, vol. 9, no. 1, p. 60, Dec. 2009. DOI: 10.1186/1471-230X-9-60.

[7] D. Regazzoni, G. de Vecchi, and C. Rizzi, “RGB cams vs RGB-d sensors: Low cost motion capture technologies performances and limitations”, *Journal of Manufacturing Systems*, vol. 33, no. 4, pp. 719–728, Oct. 1, 2014. DOI: 10.1016/j.jmsy.2014.07.011.

[8] M. Menolotto, D.-S. Komaris, S. Tedesco, B. O’Flynn, and M. Walsh, “Motion capture technology in industrial applications: A systematic review”, *Sensors*, vol. 20, no. 19, p. 5687, Oct. 5, 2020, ISSN: 1424-8220. DOI: 10.3390/s20195687.

[9] E. van der Kruk and M. M. Reijne, “Accuracy of human motion capture systems for sport applications; state-of-the-art review”, *European Journal of Sport Science*, vol. 18, no. 6, pp. 806–819, Jul. 3, 2018. DOI: 10.1080/17461391.2018.1463397.

[10] C. J. Lee and J. K. Lee, “Inertial motion capture-based wearable systems for estimation of joint kinetics: A systematic review”, *Sensors*, vol. 22, no. 7, p. 2507, Mar. 25, 2022. DOI: 10.3390/s22072507.

[11] 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, Dec. 2017. DOI: 10.1186/s12984-017-0270-x.

[12] J. Figueiredo, J. C. Moreno, A. C. Matias, F. Pereira, and C. P. Santos, "Outcome measures and motion capture systems for assessing lower limb orthosis-based interventions after stroke: A systematic review", *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 6, pp. 674–683, Aug. 18, 2021. DOI: 10.1080/17483107.2019.1695966.

[13] A. C. Alarcón-Aldana, M. Callejas-Cuervo, and A. P. L. Bo, "Upper limb physical rehabilitation using serious videogames and motion capture systems: A systematic review", *Sensors*, vol. 20, no. 21, p. 5989, Oct. 22, 2020. DOI: 10.3390/s20215989.

[14] S. D. Harbert, T. Jaiswal, L. R. Harley, T. W. Vaughn, and A. S. Baranak, "Mobile motion capture - MiMiC", in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, ISSN: 1558-4615, Jul. 2013, pp. 3435–3438. DOI: 10.1109/EMBC.2013.6610280.

[15] T. Pascu, M. White, and Z. Patoli, "Motion capture and activity tracking using smartphone-driven body sensor networks", in *Third International Conference on Innovative Computing Technology (INTECH 2013)*, Aug. 2013, pp. 456–462. DOI: 10.1109/INTECH.2013.6653712.

[16] C. Balsalobre-Fernández, G. Geiser, J. Krzyszkowski, and K. Kipp, "Validity and reliability of a computer-vision-based smartphone app for measuring barbell trajectory during the snatch", *Journal of Sports Sciences*, vol. 38, no. 6, pp. 710–716, Mar. 18, 2020, Publisher: Routledge. DOI: 10.1080/02640414.2020.1729453.

[17] J. W. L. Keogh, A. Cox, S. Anderson, B. Liew, A. Olsen, B. Schram, and J. Furness, "Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review", *PLOS ONE*, vol. 14, no. 5, e0215806, May 8, 2019, Publisher: Public Library of Science. DOI: 10.1371/journal.pone.0215806.

[18] S. Mihcin, S. Ciklacandir, M. Kocak, and A. Tosun, "Wearable motion capture system evaluation for biomechanical studies for hip joints", *Journal of Biomechanical Engineering*, vol. 143, no. 4, p. 044 504, Apr. 1, 2021. DOI: 10.1115/1.4049199.

[19] I. Klöpfer-Krämer, A. Brand, H. Wackerle, J. Müßig, I. Kröger, and P. Augat, "Gait analysis - available platforms for outcome assessment", *Injury*, vol. 51 Suppl 2, S90–S96, May 2020. DOI: 10.1016/j.injury.2019.11.011.

[20] M. Grobe-Einsler, A. Taheri Amin, J. Faber, T. Schaprian, H. Jacobi, T. Schmitz-Hübsch, A. Diallo, S. Tezenas du Montcel, and T. Klockgether, "Development of ¡span style="font-variant:small-caps;"¿ SARA $^{\text{home}}$ ¡/span¿ , a new video-based tool for the assessment of ataxia at home", *Movement Disorders*, vol. 36, no. 5, pp. 1242–1246, May 2021. DOI: `10.1002/mds.28478`. (visited on 06/03/2022).

[21] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review", *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, Feb. 2013. DOI: `10.3758/s13423-012-0333-8`.

[22] A. Baca and P. Kornfeind, "Rapid feedback systems for elite sports training", *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 70–76, Oct. 2006, Conference Name: IEEE Pervasive Computing. DOI: `10.1109/MPRV.2006.82`.

[23] C. Ernsting, S. U. Dombrowski, M. Oedekoven, J. L. O´Sullivan, M. Kanzler, A. Kuhlmey, and P. Gellert, "Using smartphones and health apps to change and manage health behaviors: A population-based survey", *Journal of Medical Internet Research*, vol. 19, no. 4, e101, Apr. 5, 2017. DOI: `10.2196/jmir.6838`.

[24] P. Krebs and D. T. Duncan, "Health app use among US mobile phone owners: A national survey", *JMIR mHealth and uHealth*, vol. 3, no. 4, e101, Nov. 4, 2015. DOI: `10.2196/mhealth.4924`.

[25] O. Oyebode, C. Ndulue, M. Alhasani, and R. Orji, "Persuasive mobile apps for health and wellness: A comparative systematic review", in *Lecture Notes in Computer Science*, Springer International Publishing, 2020, pp. 163–181. DOI: `10.1007/978-3-030-45712-9_13`.

[26] research2guidance, *Number of downloads of mHealth apps worldwide from 2013 to 2018 (in billions) [Graph]*, 2018. [Online]. Available: `https://de-statista-com/statistik/daten/studie/695434/umfrage/nummer-der-weltweiten-downloads-von-mhealth-apps/` (visited on 01/26/2022).

[27] O. Byambasuren, E. Beller, and P. Glasziou, "Current knowledge and adoption of mobile health apps among australian general practitioners: Survey study", *JMIR mHealth and uHealth*, vol. 7, no. 6, e13199, Jun. 3, 2019. DOI: `10.2196/13199`.

[28] S. Schoeppe, S. Alley, W. V. Lippevelde, N. A. Bray, S. L. Williams, M. J. Duncan, and C. Vandelanotte, "Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review", *International Journal of Behavioral Nutrition and Physical Activity*, vol. 13, no. 1, Dec. 2016, Publisher: Springer Science and Business Media LLC. DOI: `10.1186/s12966-016-0454-y`.

[29] D. Giansanti, S. Morelli, G. Maccioni, and G. Costantini, "Toward the design of a wearable system for fall-risk detection in telerehabilitation", *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, vol. 15, no. 3, pp. 296–299, Apr. 2009. DOI: `10.1089/tmj.2008.0106`.

[30] A. Sucerquia, J. D. López, and J. F. Vargas-Bonilla, "Real-life/real-time elderly fall detection with a triaxial accelerometer", *Sensors (Basel, Switzerland)*, vol. 18, no. 4, E1101, Apr. 5, 2018. DOI: 10.3390/s18041101.

[31] F. De Cillisy, F. De Simioy, F. Guidoy, R. A. Incalzi, and R. Setolay, "Fall-detection solution for mobile platforms using accelerometer and gyroscope data", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2015, pp. 3727–3730, Aug. 2015. DOI: 10.1109/EMBC.2015.7319203.

[32] J.-Y. Jenny, "Measurement of the knee flexion angle with a smartphone-application is precise and accurate", *The Journal of Arthroplasty*, vol. 28, no. 5, pp. 784–787, May 1, 2013. DOI: 10.1016/j.arth.2012.11.013.

[33] M. J. Dietz, D. Sprando, A. E. Hanselman, M. D. Regier, and B. M. Frye, "Smartphone assessment of knee flexion compared to radiographic standards", *The Knee*, vol. 24, no. 2, pp. 224–230, Mar. 1, 2017. DOI: 10.1016/j.knee.2016.11.014.

[34] L. Abou, J. Peters, E. Wong, R. Akers, M. S. Dossou, J. J. Sosnoff, and L. A. Rice, "Gait and balance assessments using smartphone applications in parkinson's disease: A systematic review", *Journal of Medical Systems*, vol. 45, no. 9, p. 87, Aug. 15, 2021. DOI: 10.1007/s10916-021-01760-5.

[35] C. Franco, A. Fleury, P. Y. Gumery, B. Diot, J. Demongeot, and N. Vuillerme, "iBalance-ABF: A smartphone-based audio-biofeedback balance system", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 211–215, Jan. 2013, Conference Name: IEEE Transactions on Biomedical Engineering. DOI: 10.1109/TBME.2012.2222640.

[36] J. Patterson A., R. Z. Amick, T. Thummar, and M. E. Rogers, "Validation of measures from the smartphone sway balance application: A pilot study", *International Journal of Sports Physical Therapy*, vol. 9, no. 2, pp. 135–139, Apr. 2014.

[37] J. A. Moral-Munoz, B. Esteban-Moreno, E. Herrera-Viedma, M. J. Cobo, and I. J. Pérez, "Smartphone applications to perform body balance assessment: A standardized review", *Journal of Medical Systems*, vol. 42, no. 7, p. 119, May 29, 2018. DOI: 10.1007/s10916-018-0970-1.

[38] *Lindera SturzApp*. [Online]. Available: https://www.lindera.de/produkte/pflege/ (visited on 06/15/2022).

[39] M. N. Kamel Boulos and S. P. Yang, "Mobile physical activity planning and tracking: A brief overview of current options and desiderata for future solutions", *mHealth*, vol. 7, pp. 13–13, Jan. 2021, ISSN: 23069740. DOI: 10.21037/mhealth.2020.01.01.

[40] S. Ishii, A. Yokokubo, M. Luimula, and G. Lopez, "ExerSense: Physical exercise recognition and counting algorithm from wearables robust to positioning", *Sensors*, vol. 21, no. 1, p. 91, Jan. 2021, Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. DOI: `10.3390/s21010091`.

[41] S. Zelman, M. Dow, T. Tabashum, T. Xiao, and M. V. Albert, "Accelerometer-based automated counting of ten exercises without exercise-specific training or tuning", *Journal of Healthcare Engineering*, vol. 2020, p. 8 869 134, 2020. DOI: `10.1155/2020/8869134`.

[42] B. C. Belyea, E. Lewis, Z. Gabor, J. Jackson, and D. L. King, "Validity and intrarater reliability of a 2-dimensional motion analysis using a handheld tablet compared with traditional 3-dimensional motion analysis", *Journal of Sport Rehabilitation*, vol. 24, no. 4, Nov. 2015, Publisher: Human Kinetics. DOI: `10.1123/jsr.2014-0194`.

[43] A. Alahmari, L. Herrington, and R. Jones, "Concurrent validity of two-dimensional video analysis of lower-extremity frontal plane of movement during multidirectional single-leg landing", *Physical Therapy in Sport*, vol. 42, pp. 40–45, Mar. 2020, Publisher: Elsevier BV. DOI: `10.1016/j.ptsp.2019.12.009`.

[44] T. J. A. Lopes, D. Ferrari, J. Ioannidis, M. Simic, F. M. d. Azevedo, and E. Pappas, "Reliability and validity of frontal plane kinematics of the trunk and lower extremity measured with 2-dimensional cameras during athletic tasks: A systematic review with meta-analysis", *Journal of Orthopaedic & Sports Physical Therapy*, vol. 48, no. 10, pp. 812–822, Oct. 2018, Publisher: Journal of Orthopaedic & Sports Physical Therapy (JOSPT). DOI: `10.2519/jospt.2018.8006`.

[45] M. Zago, M. Luzzago, T. Marangoni, M. D. Cecco, M. Tarabini, and M. Galli, "3d tracking of human motion using visual skeletonization and stereoscopic vision", *Frontiers in Bioengineering and Biotechnology*, vol. 8, Mar. 2020, Publisher: Frontiers Media SA. DOI: `10.3389/fbioe.2020.00181`.

[46] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3d human pose estimation: A review of the literature and analysis of covariates", *Computer Vision and Image Understanding*, vol. 152, pp. 1–20, Nov. 2016. DOI: `10.1016/j.cviu.2016.09.002`.

[47] *ARKit: Capturing Body Motion in 3D*. [Online]. Available: `https://developer.apple.com/documentation/arkit/content_anchors/capturing_body_motion_in_3d` (visited on 01/26/2022).

[48] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device real-time body pose tracking", 2020, Publisher: arXiv Version Number: 1. DOI: `10.48550/ARXIV.2006.10204`.

[49] X. Zhou, S. Leonardos, Xiaoyan Hu, and K. Daniilidis, "3d shape estimation from 2d landmarks: A convex relaxation approach", in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 4447–4455. DOI: `10.1109/CVPR.2015.7299074`.

[50] X. Zhou, M. Zhu, S. Leonardos, K. G. Derpanis, and K. Daniilidis, “Sparseness meets deepness: 3D human pose estimation from monocular video”, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 4966–4975. DOI: 10.1109/CVPR.2016.537.

[51] I. Akhter and M. J. Black, “Pose-conditioned joint angle limits for 3d human pose reconstruction”, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1446–1455. DOI: 10.1109/CVPR.2015.7298751.

[52] B. Martínez-Pérez, I. de la Torre-Díez, and M. López-Coronado, “Privacy and security in mobile health apps: A review and recommendations”, *Journal of Medical Systems*, vol. 39, no. 1, p. 181, Jan. 2015. DOI: 10.1007/s10916-014-0181-3.

[53] T. Dehling, F. Gao, S. Schneider, and A. Sunyaev, “Exploring the far side of mobile health: Information security and privacy of mobile health apps on iOS and android”, *JMIR mHealth and uHealth*, vol. 3, no. 1, e3672, Jan. 19, 2015, Company: JMIR mHealth and uHealth Distributor: JMIR mHealth and uHealth Institution: JMIR mHealth and uHealth Label: JMIR mHealth and uHealth Publisher: JMIR Publications Inc., Toronto, Canada. DOI: 10.2196/mhealth.3672.

[54] *CNET - Worldwide smartphone user base hits 1 billion.* [Online]. Available: https://www.cnet.com/tech/mobile/worldwide-smartphone-user-base-hits-1-billion/ (visited on 06/18/2022).

[55] L. M. Reimer, S. Weigel, F. Ehrenstorfer, M. Adikari, W. Birkle, and S. Jonas, “Mobile motion tracking for disease prevention and rehabilitation using apple ARKit”, in *Studies in Health Technology and Informatics*, IOS Press, May 7, 2021. DOI: 10.3233/SHTI210092.

[56] *Full body modeling with Plug-in Gait.* [Online]. Available: https://docs.vicon.com/display/Nexus212/Full+body+modeling+with+Plug-in+Gait (visited on 01/26/2022).

[57] L. M. Reimer, M. Kapsecker, T. Fukushima, and S. M. Jonas, “Evaluating 3D Human Motion Capture on Mobile Devices”, *Applied Sciences*, vol. 12, no. 10, p. 4806, May 10, 2022. DOI: 10.3390/app12104806.

[58] C. M. Aldenhoven, L. M. Reimer, and S. Jonas, “mBalance: Detect postural imbalance with mobile devices”, in *Studies in Health Technology and Informatics*, IOS Press, May 16, 2022. DOI: 10.3233/SHTI220344.

[59] J. Nielsen, *Usability Engineering.* Boston: Academic Press, 1993, 358 pp.

[60] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces”, in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI ’90*, Seattle, Washington, United States: ACM Press, 1990, pp. 249–256. DOI: 10.1145/97243.97281.

[61] J. Nielsen, “Evaluating the thinking-aloud technique for use by computer scientists”, in *Advances in Human-Computer Interaction (Vol. 3)*, USA: Ablex Publishing Corp., 1993, pp. 69–82.

[62] L. M. Reimer, F. Starnecker, H. Schunkert, and S. Jonas, "Developing an app for cardiovascular prevention and scientific data collection", in *Studies in Health Technology and Informatics*, IOS Press, May 7, 2021. DOI: 10.3233/SHTI210095.

[63] P. Bhatnagar, K. Wickramasinghe, E. Wilkins, and N. Townsend, "Trends in the epidemiology of cardiovascular disease in the UK", *Heart*, vol. 102, no. 24, pp. 1945–1952, Dec. 15, 2016, Publisher: BMJ Publishing Group Ltd and British Cardiovascular Society Section: Review. DOI: 10.1136/heartjnl-2016-309573.

[64] G. M. Coorey, L. Neubeck, J. Mulley, and J. Redfern, "Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data", *European Journal of Preventive Cardiology*, vol. 25, no. 5, pp. 505–521, Mar. 1, 2018. DOI: 10.1177/2047487317750913.

[65] L. Pfaeffli Dale, R. Dobson, R. Whittaker, and R. Maddison, "The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review", *European Journal of Preventive Cardiology*, vol. 23, no. 8, pp. 801–817, May 1, 2016. DOI: 10.1177/2047487315613462.

[66] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture", *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, Mar. 1, 2001. DOI: 10.1006/cviu.2000.0897.

[67] J. Aggarwal and Q. Cai, "Human motion analysis: A review", in *Proceedings IEEE Nonrigid and Articulated Motion Workshop*, Jun. 1997, pp. 90–102. DOI: 10.1109/NAMW.1997.609859.

[68] C. Cédras and M. Shah, "Motion-based recognition a survey", *Image and Vision Computing*, vol. 13, no. 2, pp. 129–155, Mar. 1, 1995. DOI: 10.1016/0262-8856(95)93154-K.

[69] D. M. Gavrila, "The visual analysis of human movement: A survey", *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, Jan. 1, 1999. DOI: 10.1006/cviu.1998.0716.

[70] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis", *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 90–126, Nov. 2006. DOI: 10.1016/j.cviu.2006.08.002. (visited on 01/21/2022).

[71] G. Sreenu and M. A. Saleem Durai, "Intelligent video surveillance: A review through deep learning techniques for crowd analysis", *Journal of Big Data*, vol. 6, no. 1, p. 48, Jun. 6, 2019. DOI: 10.1186/s40537-019-0212-5.

[72] H. Rahmalan, M. S. Nixon, and J. N. Carter, "On crowd density estimation for surveillance", pp. 540–545, Jan. 1, 2006, Publisher: IET Digital Library. DOI: 10.1049/ic:20060360.

[73] R. Ma, L. Li, W. Huang, and Q. Tian, “On pixel count based crowd density estimation for visual surveillance”, in *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, vol. 1, Dec. 2004, 170–173 vol.1. DOI: 10.1109/ICCIS.2004.1460406.

[74] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, “Recent survey on crowd density estimation and counting for visual surveillance”, *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 103–114, May 1, 2015. DOI: 10.1016/j.engappai.2015.01.007.

[75] F. Lamberti, G. Paravati, V. Gatteschi, A. Cannavò, and P. Montuschi, “Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 5, pp. 1742–1755, May 2018, Conference Name: IEEE Transactions on Visualization and Computer Graphics. DOI: 10.1109/TVCG.2017.2690433.

[76] J. K. T. Tang, J. C. P. Chan, and H. Leung, “Interactive dancing game with real-time recognition of continuous dance moves from 3D human motion capture”, in *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC ’11, New York, NY, USA: Association for Computing Machinery, Feb. 21, 2011, pp. 1–9. DOI: 10.1145/1968613.1968674. (visited on 06/08/2022).

[77] R. Okada, N. Kondoh, and B. Stenger, “A video motion capture system for interactive games”, p. 4,

[78] M. W. Whittle, “Clinical gait analysis: A review”, *Human Movement Science*, vol. 15, no. 3, pp. 369–387, Jun. 1996. DOI: 10.1016/0167-9457(96)00006-1.

[79] M. Eltoukhy, S. Asfour, C. M. Thompson, and L. Latta, “Evaluation of the performance of digital video analysis of human motion : Dartfish tracking system”, *International Journal of Scientific & Engineering Research*, vol. 3, no. 3, 2012. DOI: 10.1109/CVPR.2017.494.

[80] S. S. Paul, M. E. Lester, K. B. Foreman, and L. E. Dibble, “Validity and reliability of two-dimensional motion analysis for quantifying postural deficits in adults with and without neurological impairment”, *The Anatomical Record*, vol. 299, no. 9, pp. 1165–1173, Jul. 2016, Publisher: Wiley. DOI: 10.1002/ar.23385.

[81] S. Springer and G. Y. Seligmann, “Validity of the kinect for gait assessment: A focused review”, *Sensors*, vol. 16, no. 2, p. 194, Feb. 2016, Publisher: MDPI AG. DOI: 10.3390/s16020194.

[82] U. Puh, B. Hoehlein, and J. E. Deutsch, “Validity and reliability of the kinect for assessment of standardized transitional movements and balance”, *Physical Medicine and Rehabilitation Clinics of North America*, vol. 30, no. 2, pp. 399–422, May 2019, Publisher: Elsevier BV. DOI: 10.1016/j.pmr.2018.12.006.

[83] C. Schärer, L. v. Siebenthal, I. Lomax, M. Gross, W. Taube, and K. Hübner, “Simple assessment of height and length of flight in complex gymnastic skills: Validity and reliability of a two-dimensional video analysis method”, *Applied Sciences*, vol. 9, no. 19, p. 3975, Sep. 2019, Publisher: MDPI AG. DOI: 10.3390/app9193975.

[84] J. G. Richards, “The measurement of human motion: A comparison of commercially available systems”, *Human Movement Science*, vol. 18, no. 5, pp. 589–602, Oct. 1, 1999. DOI: 10.1016/S0167-9457(99)00023-8.

[85] *Vicon Motion Capture Systems.* [Online]. Available: https://www.vicon.com (visited on 01/26/2022).

[86] *Qualisys Motion Capture Systems.* [Online]. Available: https://www.qualisys.com (visited on 01/26/2022).

[87] 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, Feb. 1, 2005. DOI: 10.1016/j.gaitpost.2004.04.004.

[88] B. Elliott and J. Alderson, “Laboratory versus field testing in cricket bowling: A review of current and past practice in modelling techniques”, *Sports Biomechanics*, vol. 6, no. 1, pp. 99–108, Jan. 2007, Publisher: Informa UK Limited. DOI: 10.1080/14763140601058623.

[89] S. G. McLean, “Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury”, *British Journal of Sports Medicine*, vol. 39, no. 6, pp. 355–362, Jun. 2005, Publisher: BMJ. DOI: 10.1136/bjsm.2005.018598.

[90] B. Carse, B. Meadows, R. Bowers, and P. Rowe, “Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3d motion analysis system”, *Physiotherapy*, vol. 99, no. 4, pp. 347–351, Dec. 2013, Publisher: Elsevier BV. DOI: 10.1016/j.physio.2013.03.001.

[91] H. Lamine, S. Bennour, M. Laribi, L. Romdhane, and S. Zaghloul, “Evaluation of calibrated kinect gait kinematics using a vicon motion capture system”, *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 20, S111–S112, sup1 Oct. 30, 2017, Publisher: Taylor & Francis, ISSN: 1025-5842.

[92] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, Jan. 1, 2021. DOI: 10.1109/TPAMI.2019.2929257. (visited on 01/26/2022).

[93] *OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation.* [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose (visited on 01/26/2022).

[94] *Vision: Detecting Human Body Poses in Images.* [Online]. Available: https://developer.apple.com/documentation/vision/detecting_human_body_poses_in_images (visited on 01/26/2022).

[95] *TensorFlow Pose Estimate*. [Online]. Available: `https://www.tensorflow.org/lite/examples/pose_estimation/overview` (visited on 01/26/2022).

[96] *MediaPipe Pose*, https://google.github.io/mediapipe/solutions/pose.html. (visited on 03/18/2022).

[97] *Xsens Motion Capture Systems*. [Online]. Available: `https://www.xsens.com` (visited on 01/26/2022).

[98] *Perception Neuron Motion Capture*. [Online]. Available: `https://neuronmocap.com/perception-neuron-series` (visited on 01/26/2022).

[99] G. Ligorio, E. Bergamini, L. Truppa, M. Guaitolini, M. Raggi, A. Mannini, A. M. Sabatini, G. Vannozzi, and P. Garofalo, "A wearable magnetometer-free motion capture system: Innovative solutions for real-world applications", *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8844–8857, Aug. 2020, Conference Name: IEEE Sensors Journal. DOI: `10.1109/JSEN.2020.2983695`.

[100] M. P. Mavor, G. B. Ross, A. L. Clouthier, T. Karakolis, and R. B. Graham, "Validation of an IMU suit for military-based tasks", *Sensors*, vol. 20, no. 15, p. 4280, Jan. 2020, Number: 15 Publisher: Multidisciplinary Digital Publishing Institute. DOI: `10.3390/s20154280`.

[101] *Gypsy 7™ Motion Capture System*. [Online]. Available: `https://metamotion.com/gypsy/gypsy-motion-capture-system.htm` (visited on 06/10/2022).

[102] A. Stelzer, K. Pourvoyeur, and A. Fischer, "Concept and application of LPM - a novel 3-d local position measurement system", *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 12, pp. 2664–2669, Dec. 2004, Conference Name: IEEE Transactions on Microwave Theory and Techniques. DOI: `10.1109/TMTT.2004.838281`.

[103] *MotionStar Wireless®*. [Online]. Available: `http://www.mindflux.com.au/products/ascension/motionstar-wireless.html` (visited on 06/10/2022).

[104] P. Held, D. Steinhauser, A. Kamann, T. Holdgrün, I. Doric, A. Koch, and T. Brandmeier, "Radar-based analysis of pedestrian micro-doppler signatures using motion capture sensors", in *2018 IEEE Intelligent Vehicles Symposium (IV)*, ISSN: 1931-0587, Jun. 2018, pp. 787–793. DOI: `10.1109/IVS.2018.8500656`.

[105] V. C. Chen, "Detection and analysis of human motion by radar", in *2008 IEEE Radar Conference*, ISSN: 2375-5318, May 2008, pp. 1–4. DOI: `10.1109/RADAR.2008.4721059`.

[106] A.-K. Seifert, M. Grimmer, and A. M. Zoubir, "Doppler radar for the extraction of biomechanical parameters in gait analysis", *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 547–558, Feb. 2021, Conference Name: IEEE Journal of Biomedical and Health Informatics. DOI: `10.1109/JBHI.2020.2994471`.

[107] N. Golestani and M. Moghaddam, "Wearable magnetic induction-based approach toward 3d motion tracking", *Scientific Reports*, vol. 11, no. 1, p. 18 905, Sep. 23, 2021, Number: 1 Publisher: Nature Publishing Group. DOI: 10.1038/s41598-021-98346-5.

[108] T. Sato, S. Nakamura, K. Terabayashi, M. Sugimoto, and H. Hashizume, "Design and implementation of a robust and real-time ultrasonic motion-capture system", in *2011 International Conference on Indoor Positioning and Indoor Navigation*, Sep. 2011, pp. 1–6. DOI: 10.1109/IPIN.2011.6071907.

[109] K. Ashhar, C. B. Soh, and K. H. Kong, "A wearable ultrasonic sensor network for analysis of bilateral gait symmetry", in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, ISSN: 1558-4615, Jul. 2017, pp. 4455–4458. DOI: 10.1109/EMBC.2017.8037845.

[110] D. Laurijssen, S. Truijen, W. Saeys, and J. Steckel, "Three sources, three receivers, six degrees of freedom: An ultrasonic sensor for pose estimation & motion capture", in *2015 IEEE SENSORS*, Nov. 2015, pp. 1–4. DOI: 10.1109/ICSENS.2015.7370689.

[111] Y. Qi, C. B. Soh, E. Gunawan, and K.-S. Low, "A wearable wireless ultrasonic sensor network for human arm motion tracking", in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, ISSN: 1558-4615, Aug. 2014, pp. 5960–5963. DOI: 10.1109/EMBC.2014.6944986.

[112] *Apple iPhone*. [Online]. Available: https://www.apple.com/iphone/ (visited on 06/18/2022).

[113] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3d markerless motion capture accuracy using OpenPose with multiple video cameras", *Frontiers in Sports and Active Living*, vol. 2, May 2020, Publisher: Frontiers Media SA. DOI: 10.3389/fspor.2020.00050.

[114] E. D'Antonio, J. Taborri, E. Palermo, S. Rossi, and F. Patane, "A markerless system for gait analysis based on OpenPose library", in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE, May 2020. DOI: 10.1109/i2mtc43012.2020.9128918.

[115] M. Silverio-Fernández, S. Renukappa, and S. Suresh, "What is a smart device? - a conceptualisation within the paradigm of the internet of things", *Visualization in Engineering*, vol. 6, no. 1, p. 3, May 9, 2018. DOI: 10.1186/s40327-018-0063-8.

[116] B. L. Risteska Stojkoska and K. V. Trivodaliev, "A review of internet of things for smart home: Challenges and solutions", *Journal of Cleaner Production*, vol. 140, pp. 1454–1464, Jan. 1, 2017. DOI: 10.1016/j.jclepro.2016.10.006.

[117] R. Lanotte and M. Merro, "A semantic theory of the internet of things", *Information and Computation*, vol. 259, pp. 72–101, Apr. 1, 2018. DOI: 10.1016/j.ic.2018.01.001.

[118] I. Bisio, C. Garibotto, F. Lavagetto, A. Sciarrone, and S. Zappatore, "Unauthorized amateur UAV detection based on WiFi statistical fingerprint analysis", *IEEE Communications Magazine*, vol. 56, no. 4, pp. 106–111, Apr. 2018, Conference Name: IEEE Communications Magazine. DOI: 10.1109/MCOM.2018.1700340.

[119] W. Sousa, E. Souto, J. Rodrigres, P. Sadarc, R. Jalali, and K. El-Khatib, "A comparative analysis of the impact of features on human activity recognition with smartphone sensors", in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, ser. WebMedia '17, New York, NY, USA: Association for Computing Machinery, Oct. 17, 2017, pp. 397–404. DOI: 10.1145/3126858.3126859.

[120] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks", *Expert Systems with Applications*, vol. 59, pp. 235–244, Oct. 15, 2016. DOI: 10.1016/j.eswa.2016.04.032.

[121] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition", *IEEE Access*, vol. 5, pp. 3095–3110, 2017, Conference Name: IEEE Access. DOI: 10.1109/ACCESS.2017.2676168.

[122] D. Mukherjee, R. Mondal, P. K. Singh, R. Sarkar, and D. Bhattacharjee, "EnsemConvNet: A deep learning approach for human activity recognition using smartphone sensors for healthcare applications", *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31 663–31 690, Nov. 1, 2020. DOI: 10.1007/s11042-020-09537-7.

[123] F. Ma, G. V. Cavalheiro, and S. Karaman, "Self-supervised sparse-to-dense: Self-supervised depth completion from LiDAR and monocular camera", in *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada: IEEE, May 2019, pp. 3288–3295. DOI: 10.1109/ICRA.2019.8793637.

[124] T. X. Tran, C.-k. Kang, and S. L. Mathis, "Lower-gait tracking mobile application: A case study of lower body motion capture comparison between vicon t40 system and apple augmented reality", in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2020, pp. 2654–2656. DOI: 10.1109/BIBM49941.2020.9313201.

[125] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields", in *CVPR*, 2017. DOI: 10.1109/TPAMI.2019.2929257.

[126] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines", in *CVPR*, 2016. DOI: 10.1109/CVPR.2016.511.

[127] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping", in *CVPR*, 2017.

[128] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards accurate multi-person pose estimation in the wild", presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4903–4911.

[129] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, and K. Murphy, "PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model", arXiv, arXiv:1803.08225, Mar. 22, 2018, type: article. DOI: `10.48550/arXiv.1803.08225`. arXiv: `1803.08225[cs]`.

[130] L. Xie and X. Guo, "Object detection and analysis of human body postures based on TensorFlow", in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Aug. 2019, pp. 397–401. DOI: `10.1109/SmartIoT.2019.00070`.

[131] S. E. Adi and A. J. Casson, "Design and optimization of a TensorFlow lite deep learning neural network for human activity recognition on a smartphone", in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, ISSN: 2694-0604, Nov. 2021, pp. 7028–7031. DOI: `10.1109/EMBC46164.2021.9629549`.

[132] *Counting Human Body Action Repetitions in a Live Video.* [Online]. Available: `https://developer.apple.com/documentation/createmlcomponents/counting_human_body_action_repetitions_in_a_live_video_feed` (visited on 06/10/2022).

[133] T. Yan, Z. Hu, Y. Qian, Z. Qiao, and L. Zhang, "3d shape reconstruction from multifocus image fusion using a multidirectional modified laplacian operator", *Pattern Recognition*, vol. 98, p. 107 065, Feb. 1, 2020. DOI: `10.1016/j.patcog.2019.107065`. (visited on 06/20/2022).

[134] Z. Zhang, "Analysis of the advantages of the M1 CPU and its impact on the future development of Apple", in *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, Zhuhai, China: IEEE, Sep. 2021, pp. 732–735. DOI: `10.1109/ICBASE53849.2021.00143`.

[135] *HerzFit Apple App Store.* [Online]. Available: `https://apps.apple.com/de/app/herzfit/id1618789597` (visited on 06/20/2022).

[136] *HerzFit Google Play.* [Online]. Available: `https://play.google.com/store/apps/details?id=app.herzfit` (visited on 06/20/2022).

[137] *Mobile Operating System Market Share Worldwide.* [Online]. Available: `https://gs.statcounter.com/os-market-share/mobile/worldwide/#monthly-202105-202205` (visited on 06/21/2022).

[138] R. Green and L. Guan, "Quantifying and recognizing human movement patterns from monocular video images-part II: Applications to biometrics", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 191–198, Feb. 2004, Conference Name: IEEE Transactions on Circuits and Systems for Video Technology. DOI: `10.1109/TCSVT.2003.821977`.