UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics and Computer Science

The Development of a Novel Method for Measuring Biomechanical Features of Agility Using Markerless Motion Capture

Jasper Jochem Peetsma
MSc. Interaction Technology
December 2023

Graduation Committee
Dr. D.B.W. Postma, Human Media Interaction
Dr. Ir. D. Reidsma, Human Media Interaction
Dr. Ir. B.J.F. Van Beijnum, Biomedical Signals and Systems
Ing. M.H.H. Weusthof, Biomedical Signals and Systems

Human Media Interaction Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
Enschede, The Netherlands

Abstract

Assessing the agility of sports practitioners up until now has been done through agility tests and change-of-direction tests, in which the test completion time is the sole metric of performance. Within this thesis, a novel method is proposed to better understand and estimate an athlete's agility. As time is often of the essence – especially for professional sports players and teams – the method takes practicality into account by capturing movement without needing markers or sensors to be attached to the athlete's body. This markerless motion capture (MMC) method saves time, as only a set of two cameras needs to be set up in advance, and no extra valuable time is wasted in between the data collection of each athlete. Literature research showed correlations between agility and inter-limb asymmetry, ground contact time, and take-off distance. To measure these features, OpenPose was used to detect pose keypoints in the stereo video recordings of arrowhead agility tests, which were triangulated into 3D coordinates using MATLAB. These coordinates were further developed into parameters, or key performance indicators, by calculating (differences in) joint angles, ground contact start and end times, and whole-body centre-of-mass positions, velocity and acceleration. To demonstrate how the data can be used, parameters are modelled through a generalized linear mixed-effect modelling (GLMM) approach. However, no definitive conclusions can be drawn from the model without validating the data with a gold-standard motion capture system. The work shows promising results for the use of OpenPose in the assessment of agility. The system offers many possibilities for easy collection of biomechanical data using off-the-shelf tools, even outside of lab environments. However, there is room for improvement in several steps of the process (e.g. in the camera calibration, the OpenPose configuration, the triangulation into 3D coordinates, and the joint angle calculation). These must be addressed before accurate and reliable information can be extracted.

Acknowledgements

First and foremost, I would like to thank my supervisor, dr. Dees Postma, for his continuous support and enthusiasm, and the critical eye he lent me every week for the past months, sharing his knowledge of movement sciences, interaction technology, and the research process itself. It kept me engaged, enthusiastic and motivated to continue working throughout the process. I am thankful for having had such a nice and down-to-earth supervisor. If I ever hear a fellow I-Tech student searching for a supervisor, I know who I'd recommend.

Secondly, I would like to thank everyone who has helped me during the process by sharing their knowledge, giving me new insights, sharing their enthusiasm for my project with me, and not treating me like just another student but as an equal. It was very nice to be included in a project spanning multiple organisations with different knowledge and backgrounds.

Third, I would like to thank the FC Twente/Heracles Academie for organising the two days full of testing and measuring and allowing me to collect data there as well. Additionally, I would like to thank my fellow students and, once again, my supervisor, who voluntarily helped me in correctly recording the data of close to 200 tests. Furthermore, I would like to thank all participants who allowed themselves to be recorded for the sake of science.

Lastly, I would like to thank my friends and family for their support. It really helped me to study together to stay motivated and remain focused, especially in the stage of writing the report. At the same time, they also provided me with distraction every now and then, which was sometimes much-needed.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Agility is a complex human quality that is especially essential in team sports (Paul et al., 2016). Specifically in invasion sports, which typically include a substantial amount of interaction between attackers and defenders, team performance highly depends on the individual athletes' agility (Young et al., 2022). Agility comprises a reactive, cognitive component as well as a physical element. It can be defined as "a rapid whole-body movement with change of velocity or direction in response to a stimulus" (Sheppard et al., 2006). In the past, it has been difficult to measure the whole of agility, as most "agility tests" are, in fact, so-called change-of-direction (COD) tests. They merely assess the physical aspect of agility (i.e. how well a person can decelerate, accelerate, and change direction), leaving out any conscious cognitive activities. The tests are most often measured by the time it takes the participant to complete the test, which is referred to as the change-of-direction speed (CODS). Common tests that measure COD in terms of time are the t-test (Semenick, 1990), the 5-0-5 test (Draper, 1985), the pro-agility (shuttle) test (Forster et al., 2022; McKay et al., 2020; Nimphius et al., 2016), and the arrowhead test (Rago et al., 2020). Additionally, FC Twente (n.d.), a professional football club from Enschede, has developed a combination of the arrowhead and 5-0-5 test to routinely check in on their players' CODS. A problem with all of these COD tests is that they lack explanatory power regarding underlying factors. These are perhaps related to kinematic and kinetic (i.e. biomechanical) qualities, techniques and strategies, such as initial take-off, foot placement, or control of their centre-of-mass position. However, currently, such metrics are not available to sports practitioners during training on the field.

Moreover, especially in professional sports, practice time is considered a valuable resource that is not to be wasted. In the same preliminary expert interviews that showcased this, it was found that the football club at hand, FC Twente, sees the actual practice of football as the most effective method of improving performance. This is because the ecological validity of measuring agility in a sport-specific scenario is higher than in separate tests not resembling that specific context. Therefore, agility is ideally measured during regular football training, not "wasting" time taking separate tests. However, as it is still unknown what should be measured, this is not yet feasible. Therefore, the discussion brings forth the need for a method of measuring biomechanical features of agility in a practical, non-time-consuming manner.

A promising technology worth investigating to find a solution to the described problem is found in markerless motion capture (MMC). This type of motion capture uses camera footage to detect body poses and has no need for on-body markers or sensors (see section 2.3 for an elaboration on the topic). This means that setting up the equipment would only have to be done before the test session, and the extra time taken to measure the COD factors would be virtually zero. Additionally, MMC might also make it possible to analyse tests and training retrospectively. However, no implementations for measuring agility were found for this relatively new computer vision technology. Therefore, the question arises whether the fast-paced movements typically associated with agility and COD tests can be accurately captured. Generally, a marker- or sensor-based approach is more accurate, but it is also more time-consuming as markers or sensors must be attached carefully to each participant before they can perform a COD test.

1.1 Research Questions

Given the observed problems described above, the research conducted in this thesis will be split into two research questions. First, the knowledge gap must be closed relating to the biomechanical features of agility beyond the metric of time and how the found features can be measured. The first research question to address this is formulated as follows:

> **RQ1: How can biomechanical features of agility be measured in a practical manner beyond the confines of a lab setting?**

Through the development of a method implementing the findings of RQ1, the applicability of the markerless motion capture method can be evaluated. This will answer the following question:

> **RQ2: To what extent is markerless motion capture suitable for measuring biomechanical features of agility?**

1.2 Thesis Outline

This thesis will cover the main elements that are required to answer the research questions. In chapter 2, related work is outlined concerning existing technology and research relevant to the topic of this thesis. The chapter concludes with the design choices made based on the exploration of the current state of the art. The following chapter 3 encompasses the full methodology, from the initial study design to the method used for data analysis. The results and their interpretation and implications, as well as the limitations of this study, are discussed in chapter 4. Both research questions are answered in this chapter. Finally, chapter 5 summarizes the main findings of the research conducted and recommends future directions for research.

Chapter 2

Related Work

2.1 Agility and Change-Of-Direction Speed

Before any methods of measuring agility can be investigated, it must be known what agility is. For a long time, no clear definition existed for agility. In a review on the topic, Sheppard et al. (2006) found that the term has had different meanings in research. They report that agility has been referred to as the ability to change direction rapidly and, in some cases, accurately. Others add a change of direction of the whole body or a rapid change in movement and direction of the limbs. It is also argued that the action should be in response to, for example, "another patient's movement, movement of the opponent, movement of play, or movement of the ball" (Welling et al., 2021). This is because agility is believed to have a cognitive component (see figure 1) that cannot be trained if the trainee already knows what will happen. Naturally, decision-making skills cannot be trained if the decision is made before the training. In that case, only the physical components of agility would be trained.

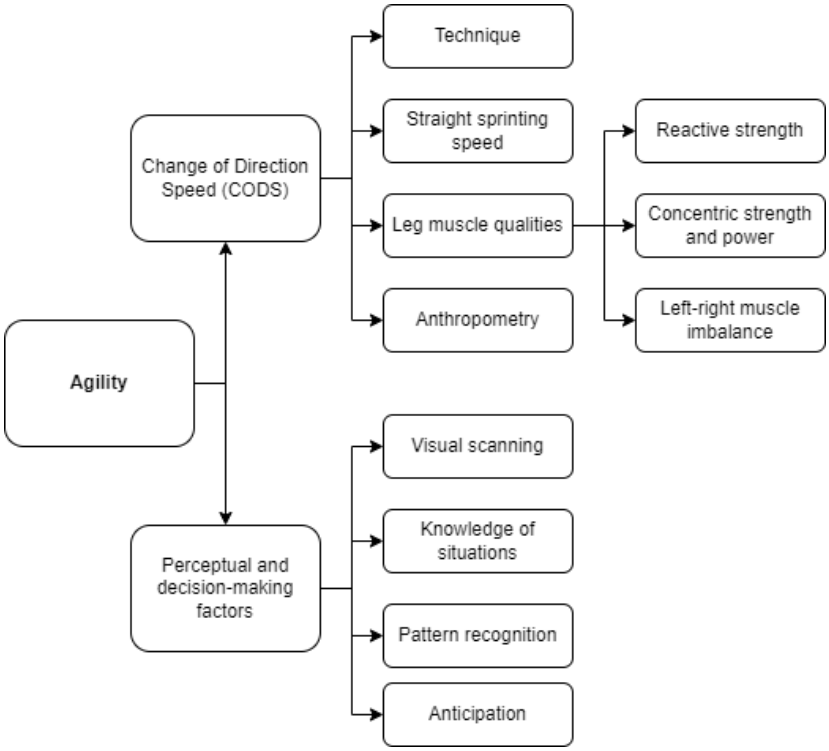


Figure 1: Components of agility (modified from Sheppard et al. (2006)).

Therefore, two types of agility can be distinguished: preplanned and reactive agility (Šimonek et al., 2016). In preplanned agility, it is known what to do beforehand, while in reactive agility, participants react to some stimulus. It can be argued whether preplanned agility should even be called agility due to the lack of such a stimulus and whether it should be called change-of-direction speed (CODS) instead. Likewise, it is argued that the precursing "reactive" is redundant (Young et al., 2015). There-

fore, for the remainder of this project, the definition of agility (for sports) that will be used is the one from Sheppard et al. (2006)'s review: "a rapid whole-body movement with change of velocity or direction in response to a stimulus". Thus, movements that do not include the reaction to a stimulus are regarded as (preplanned) change-of-direction movements.

CODS is an element of agility, yet Young et al. (2015) state that agility and CODS are specific skills independent of one another. Although there is most likely a difference in how an athlete cuts around a corner between a test in which the direction to cut to is known beforehand (i.e. a COD test) compared to the same test where this is not known (i.e. an agility test), it is expected that the types of movement are similar. Therefore, as the cognitive component does not lie within the scope of this thesis, the next section will dive into existing methods of measuring CODS.

2.2 Measuring Change-Of-Direction Speed

This thesis aims to find a method of measuring (the CODS part of) agility through the novel use of markerless motion capture. It is important to know existing methods using other resources, which will be explored in this section.

2.2.1 COD Tests in Lab Environments

Most of the existing methods for measuring either agility or CODS employ a lab setting where participants move through a specified parkour as quickly as possible. Therefore, the sole measure used in these methods is time: the faster the participant is, the more agile they are considered to be. One example of such a test procedure is the t-test, which is displayed in figure 2a (Semenick, 1990). In this test, the participant sprints from point A to B, then shuffles sideways to point C, then shuffles to point D and back to B, after which they run backwards to point A. It measures the participant's ability to "change directions rapidly while maintaining balance without loss of speed".

Another popular test is the 5-0-5 test, which, as opposed to the t-test, only requires the participant to run forward (Draper, 1985). Its set-up is shown in figure 2b. Here, players run from point A through B – where a gate is triggered to start timing – to the line at point C, where they turn around and run through the timing gates at point B again and do not stop running until they pass point A again. This test is particularly fitting for cricket players, but "the test has [also] been used for other sports requiring change of directions and agility" (Sheppard et al., 2006).

A test that is slightly more complicated is the arrowhead test, which is proven to be reliable in measuring CODS in football players (Rago et al., 2020). Displayed in figure 2c, the player starts behind the timing gates at A and runs around either B_L or B_R, followed by turning around point C_L or C_R, respectively, then around D, past B and back to A through the timing gates. As opposed to the other tests displayed in figure 2, the arrowhead test requires 2-meter high poles instead of small pylons. The reason for this is that it is required to run around the poles with the whole body, not just with the feet. The test has been shown to have "very high validity and reliability and can be used by sports coaches to evaluate the training process" (Chalil et al., 2017).

Similar to the 5-0-5 test is the pro-agility test, whose name might cause some confusion as it is essentially a change-of-direction test (McKay et al., 2020; Nimphius et al., 2016). In this test (see figure 2d), the participant starts behind the timing gates at point A, runs to B, where they turn around and run past A to C. Here, they turn around once more and run through the timing gates at A to stop the time. The pro-agility test has been used to determine the change-of-direction speed in athletes of several team sports as the movements resemble the transitions made when shifting from attacking to defending and vice versa (Forster et al., 2022).

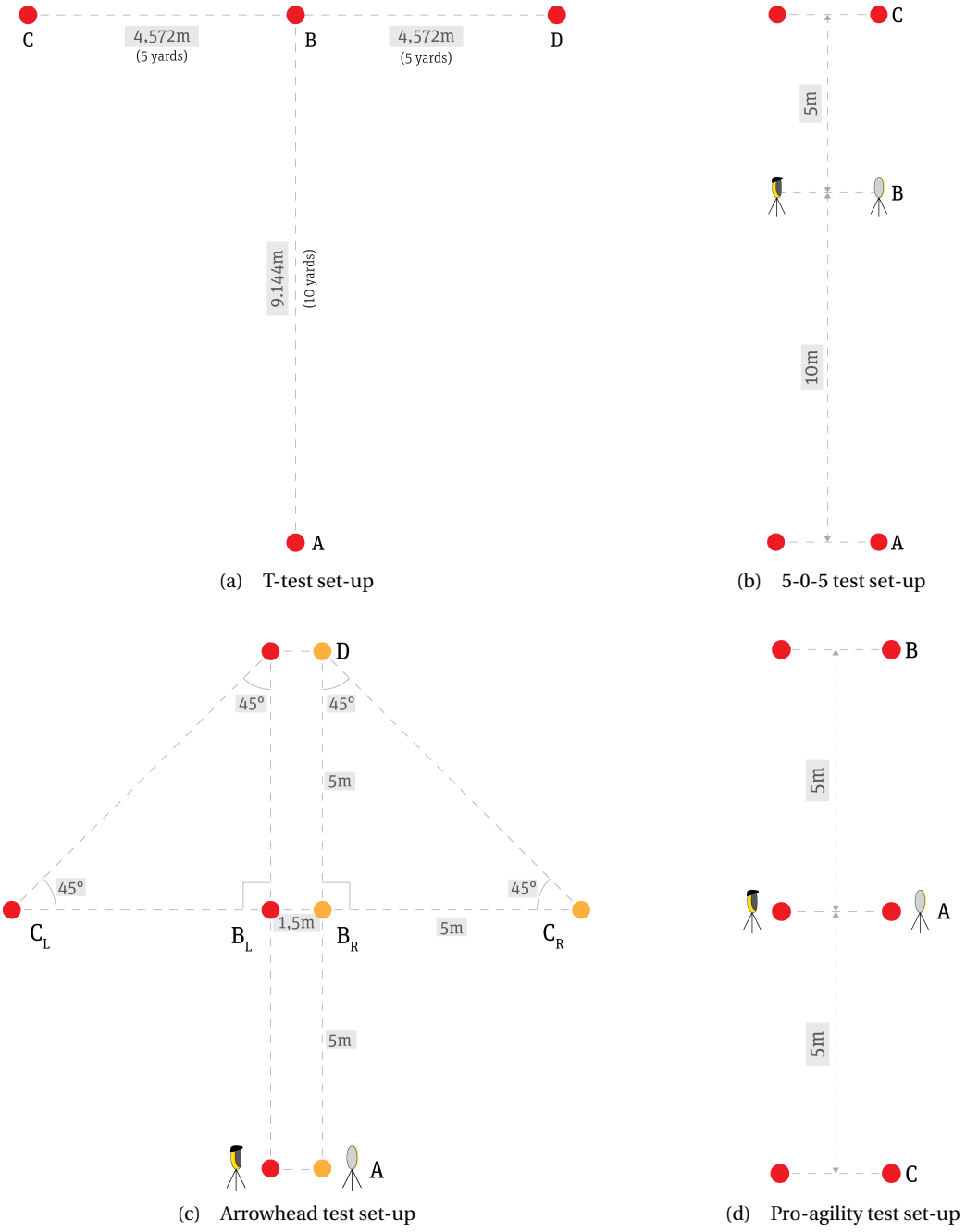


Figure 2: Test set-ups of several change-of-direction tests.

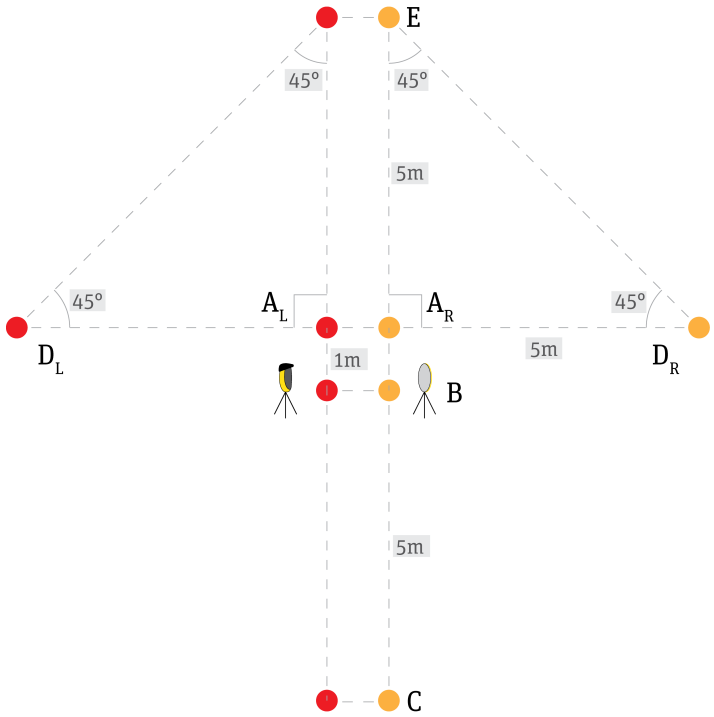


Figure 3: Test set-up for the "Agility FC Twente" change-of-direction test.

An additional change-of-direction test has been developed by FC Twente and its joint academy with Heracles Almelo, which is a combination of the arrowhead test and the 5-0-5 test. The test does not have an official name but will be referred to as the "Agility FC Twente" test, or AFCT for short. The test set-up is displayed in figure 3. In this change-of-direction test, the participant starts at point A and triggers the timing gates by running past point B to point C. Here, they turn around, run past point B again and around either point A_L or A_R, then around point D in the same direction, around point E, and finally, run past A and B and only stop running once they are past point C. The AFCT test is not yet validated. This would require studying its ability to judge a participant's level of agility by, for example, checking the extent to which participants' performances on the test correlate to those on a different, validated test.

2.2.2 Exercises Prior to COD Tests

While all of the above tests can be used to assess COD performance, studies exist that have investigated some of these COD tests one step further by adding a second exercise prior to the routine. One study investigated whether exercises designed to augment vertical stiffness influenced CODS (Maloney et al., 2019). They describe vertical stiffness as "the vertical displacement of the center of mass in response to vertical ground reaction force during sagittal plane movement, [which] seeks to approximate the deformation of the leg-spring at instants of ground contact". They compared the effect of unilateral (i.e. using only one lower limb) and bilateral (i.e. using both lower limbs) interventions conducted prior to a COD test. The study concluded that a significant improvement in CODS was observed only for unilateral exercises. Therefore, they recommend incorporating unilateral exercises designed to augment vertical stiffness in training for sports where CODS is relevant.

Similarly, Dann et al. (2022) investigated how the plyometric exercise (i.e. including an explosive element) called alternate leg bounding might influence performance. In this exercise, the subject runs in a fashion in which they try to cover as much distance as possible with each step (called a bound) while maximising the height of the knees, heels and toes (Trackwired, n.d.). The experiment was done on a grass surface by one group and on a hard surface by another. All participants were female team sports players. The subjects performed three consecutive 20-meter pro-agility tests 4, 8, and 12 minutes after the bounding intervention. Results indicate that when 8 to 12 min-

utes of recovery time is given, CODS is improved significantly. This suggests that, independently of the training surface, alternate leg bounding can be used to enhance team sports players' CODS performance.

2.2.3 Time-Independent COD Measurements and Correlations

All COD tests described in section 2.2.1 (except for the AFCT test) are validated tests, meaning they are proven to be able to measure CODS. However, they rely purely on the metric of time. Some studies have proposed other metrics that go beyond only measuring the completion time of a test. These might be useful for a more technical or biomechanical assessment of agility, potentially measurable with MMC.

Sprint speed is an inherent part of the COD tests mentioned earlier. To isolate COD ability independent of sprint speed, Nimphius et al. (2016) introduced a measure called the change of direction deficit. The researchers had 17 cricketers perform 5-0-5 tests in which the foot of their final ground contact before turning was instructed: three times turning off the right foot and three times off the left foot (in random order). Additionally, they performed a 30-meter sprint, measured at the 10m and 30m marks. The difference between the average times of the two tests was calculated as the COD deficit, which was proven to be significant. However, the metric can only be measured by having athletes perform multiple tests in lab environments. This is not desired, as it takes extra time during training. However, when these tests are already performed as part of a medical and physical screening, then it would be simple to calculate this metric.

The practicality criterion could be satisfied by another study. Philipp et al. (2021) investigated whether a difference in the performance of the left and right limbs could cause a different performance in COD tests. They confirmed that interlimb asymmetries in mostly mean peak velocity (mPV) and mean peak power (mPP) could indeed be used to discriminate between faster and slower COD test performances. Similarly, Bishop et al. (2021) concluded that larger asymmetries could be associated with reduced performance in CODS. This suggests that interlimb asymmetry might be useful as a Key Performance Indicator (KPI) for CODS. A way to measure this exists, but it does require a lab environment. The limb symmetry index (LSI) is the ratio between performances of a hop test (e.g. the single hop or triple hop) using the left and right leg (Gokeler et al., 2017). In rehabilitation, an LSI of 90% is often used as a threshold for patients to allow them to return to their preferred sports practice again. However, symmetry can also be gauged by comparing other features measured from the left and right limbs. The range of motion (ROM), for example, has been used as a "basic marker for clinical outcome studies" (Lea et al., 1995). Measurements of joint motion have functioned as a way to "assess injuries and diseases in the locomotor system" (Roaas et al., 1982), and its range is generally more restricted in older people.

In contrast to these findings, however, Thomas et al. (2020) found that asymmetry does not influence completion time. They found that it does, however, influence ground reaction forces (GRF) and knee abduction angles (KAA) (Englander et al., 2019). A non-straight KAA is commonly referred to as bow legs, where the knees are pushed away from the centre of the body. Findings of Thomas et al. (2020) suggest that higher asymmetry may lead to higher horizontal GRF during the final ground contact before pivoting with the non-dominant limb. Thus, when a subject had shown to have high interlimb asymmetry, in a COD test involving a 180°turn, they would produce greater braking during their last ground contact when instructed to turn with their dominant foot compared to turning with their non-dominant foot. This suggests that the technique used for changing direction within COD tests is inconsistent and is dependent on pivoting direction and foot preference. As for the knee abduction angles, the authors found significant differences between the dominant and non-dominant limbs: "Turning off the [dominant] limb showed increased KAA; thus, coaches and practitioners are encouraged to coach a 180° COD strategy, which emphasizes loading in the sagittal plane (hip, knee, and ankle alignment) and limits a “knee valgus” position". The latter can be described by a position in which the knee is bent slightly in the opposite direction as compared to "normal" bending of the knee, which is linked to an increased potential for injury. Additionally, the extent to which the knees are pushed outward is known as the knee abduction moment, which is associated with increased vertical centre of mass deviations and knee injury risk (Bill et al., 2022).

The role of trunk control during an agility test (i.e. including the response to a stimulus) was explored by Edwards et al. (2017). They divided participants based on their level of trunk control, in

which high trunk control meant participants demonstrated a high range of motion of the trunk relative to the pelvis. Participants with high trunk control showed better performance in an agility test (Getchell, 1979), displayed "higher [counter-movement jump] height, lower knee flexion angles, greater trunk lateral flexion and rotation relative to [the] pelvis, and greater angular momentum" (Edwards et al., 2017).

A quite different approach was employed by Welch et al. (2019), who analysed the correlation between joint-based biomechanical variables in cutting (i.e. a form of changing direction) and cutting performance using a principal component analysis (PCA). A 110° and a 45° cut were investigated, and a total of 200 variables related to angles, angular velocities and moments were considered for axes in x, y, and z direction and for several events and phases of the cutting movement. Principal components (PC) were identified for both cuts, existing of multiple variables with high correlation to performance. The identified PCs shared three performance cues: increasing the distance between the centre of mass and foot placement during the eccentric phase (i.e. while decelerating before turning), using shorter ground contact time, and maintaining a low centre of mass during the concentric phase (i.e. while re-accelerating after turning). A study by Hewit et al. (2012) resulted in a similar finding regarding the eccentric phase, as they report the "distance from the foot of the trail leg to the center of mass" should be kept large. Dos'Santos et al. (2017) and Condello et al. (2016) confirmed that a shorter ground contact time is beneficial to COD performance. Furthermore, Dos'Santos et al. (2017) found that faster athletes showed significantly greater horizontal propulsive forces, greater horizontal braking forces in the second to last ground contact (especially when compared to the final ground contact), and lower vertical impact forces.

2.2.4 Concluding Remarks

Throughout this section, it has become clear that CODS can be assessed using various tests. Studies have shown it is possible to improve performance on such tests by having participants perform an additional exercise before taking the COD test. However, these require extra time from the participant. Therefore, other correlations were sought after that could be measured during a COD test besides time. It came to light that people who are relatively asymmetrical in their limbs seem to underperform as compared to people who are relatively symmetrical. While the level of symmetry has been assessed using left and right hop tests, it is expected that differences in the range of motion might exhibit a similar influence on CODS. Similarly, a higher trunk range of motion relative to the pelvis has been associated with increased agility. Furthermore, research revealed that improved CODS might be achieved by keeping a large distance between trail leg foot placement and the centre of mass, a low centre of mass, and a short ground contact time. Lastly, agility has been associated with greater horizontal propulsive and braking forces and lower vertical impact forces. These findings were used to determine the features that will be measured using markerless motion capture.

2.3 Markerless Motion Capture (MMC)

Seeing that an understanding has been established of what agility and CODS are and how they can be measured, the existing options for markerless motion capture can be explored. The MMC method is required to be accurate and practical in use so that the data reflects reality in a meaningful way and the data is easily collectable. As there is a multitude of MMC software packages that can be used, they are discussed and considered in this section.

2.3.1 Closed-Source and Commercially Available MMC

Some companies exist that offer their MMC software for a price. One of these closed-source systems that has shown great potential is Theia3D (Theia Markerless Inc., n.d.), which is offered as a software package or as a full package, including all required hardware. The company states that it "distinguishes itself from the other markerless tracking solutions through its ability to provide a highly accurate and generalized solution to markerless motion capture". Additionally, they declare that every aspect "has been selected to ensure biomechanical accuracy and relevance of the results and to follow standard practice and conventions in the field". To use their software, a demo can be booked in which the customer must specify the industry they work in, as well as the camera system

they intend to use. This way, they want to tailor the package they offer to the specific needs of the customer. However, this restricts Theia3D for use in research.

Another MMC system that might be better known than Theia3D is Kinect. This commercially available system (now discontinued) was originally built to allow people to play games using their body as the controller. The system consists of "a depth sensor, a colour camera, and a four-microphone array that provide full-body 3D motion capture, facial recognition, and voice recognition capabilities"(Zhang, 2012). The technology has been widely used in a variety of research areas, such as hand-gesture recognition, human-activity recognition, body biometrics estimation (such as weight, gender, or height), 3D surface reconstruction, and healthcare applications (Zhang, 2012). As Kinect is discontinued, the hard- and software are no longer updated, meaning the system would not be a future-proof option.

2.3.2 Open-Source MMC

As opposed to the paid closed-source software packages, open-source software is free to use and available to everyone. It can be argued that these advantages make this type of software more interesting for research purposes. One example of open-source MMC software was presented by Google Research through BlazePose: a "lightweight convolutional neural network architecture for human pose estimation that is tailored for real-time inference on mobile devices" (Bazarevsky et al., 2020). BlazePose employs a top-down paradigm, which first detects the person in an image and then performs pose estimation in a smaller region of the image (Geng et al., 2021). Due to its lightweight nature, it is suitable for real-time use. However, BlazePose is limited to single-person human pose detection. The software has also been shown to have a relatively low keypoint detection rate, meaning that it is less capable of consistently detecting keypoints in an input video (Mundt et al., 2023).

A similar open-source tool employing the same top-down paradigm is AlphaPose (Fang et al., 2023). In contrast to BlazePose, AlphaPose is able to detect multiple persons in an image. AlphaPose "first detects [each] person and then performs single-person pose estimation for each detected person" (Geng et al., 2021). The reverse can be referred to as a bottom-up paradigm, where keypoints are detected first and then associated with a person (Geng et al., 2021). An open-source multi-person pose estimation tool employing this paradigm is OpenPose (Cao et al., 2021). The system uses part affinity fields (PAF) "to learn to associate body parts with individuals in the image" (Cao et al., 2021). A PAF is a set of 2D vector fields that encode the location and orientation of limbs over the image domain (Cao et al., 2016), which are used to associate keypoints with limbs and, eventually, with the full body. A bottom-up approach like OpenPose is more robust to occlusion and complex poses (Jin et al., 2017). Generally, a top-down approach is more accurate, but a bottom-up approach is more efficient (Geng et al., 2021). In an evaluation conducted by Mundt et al. (2023), OpenPose and AlphaPose both achieve higher detection rates than BlazePose. Additionally, BlazePose has been found to deviate from anatomical joint centres more often than AlphaPose and OpenPose, resulting in less accurate joint angle calculations (Mroz et al., 2021).

A slightly different approach from the previously mentioned pose detection methods is employed by ViTPose (Xu et al., 2022). This method uses "simple and non-hierarchical vision transformers as backbones to generate feature maps for the given human instances" (Lovanshi et al., 2022). Despite the method being fairly simple, Xu et al. (2022) report "surprisingly good capabilities of ViTPose [regarding] simplicity, scalability, flexibility, and transferability". ViTPose might offer a good solution for scenarios that require low computational costs.

2.4 Biomechanical Analysis Using MMC

Whichever markerless motion capture system is used within this thesis, all can be used to analyse biomechanical features of humans such as joint angles and motion. However, the specific MMC methods used, as well as the results, differ in every study. It is important to understand what methods have been investigated in the past and what methods have proven successful in related implementations of MMC. It was found that common questions asked by researchers relate to the validity and reliability of MMC in acquiring gait parameters (Kanko et al., 2021b; Riazati et al., 2022; Sandau et al., 2014). Theia3D has shown great potential, as studies have demonstrated its reliability for biomechanical and clinical use (Kanko et al., 2021a,b; Riazati et al., 2022). A study on the Kinect V2

has also shown promising results. However, it is explicitly mentioned that "appropriate correction procedures" must be performed (Tanaka et al., 2018), suggesting that the raw data acquired from the system might not be sufficiently accurate. Furthermore, Sandau et al. (2014) did not use an off-the-shelf tool but developed a method in which they generate a 3D point cloud based on surface textures of around 75,000 3D points from an 8-camera set-up. Results showed that although "flexion/extension angles, as well as hip abduction/adduction, closely resembled those obtained from the marker-based system, [...] the internal/external rotations, knee abduction/adduction and ankle inversion/eversion were less reliable"(Sandau et al., 2014).

Another study focused on a different biomechanical feature: the ground contact time. Calculating the feature during running has been proven feasible using accelerometers through the research of Purcell et al. (2005). Their approach provided very close estimates of actual contact time during running. Keypoints retrieved from MMC provide information that is rather similar to accelerometer data. While accelerometers measure acceleration (in m/s^2), MMC data can be processed to acquire positional data over time (i.e. velocity), of which the derivative is acceleration. This idea is implemented for human gait analysis by Stenum et al. (2021). They call the moment the heel first touches the ground the heel strike and the moment the toes leave the ground the toe-off. These events are defined by "the time points of positive and negative peaks of the anterior-posterior ankle trajectories relative to the pelvis"(Stenum et al., 2021). These two events make way for a multitude of spatiotemporal gait parameters, namely the step time, stance time, swing time, double support time, step length, and gait speed. Differences were found while comparing the markerless OpenPose-based parameters with the same parameters calculated with marker-based motion capture. However, they state that their workflow is accurate enough for the purpose of their study, which was to detect changes in the gait pattern.

Yet another study implemented MMC to measure vertical jump height based on an approximation of the body's centre of mass. Webering et al. (2021) compare this to reference heights retrieved using an 8-camera Vicon set-up (i.e. a benchmark marker-based motion capture system). The height measurement is based on a marker attached to the lower back of the participant. They conclude that the method using only OpenPose increases the ease of use as compared to a marker-based method, but the jump height measurements are slightly less accurate.

2.5 Summary of The Related Work

In the previous sections, related work has been discussed regarding the definition and measurement of agility, the available markerless motion capture options, and existing implementations of MMC. This information is used to guide the following steps of the project in order to answer the research questions.

Change of direction has been measured in several ways in the past, but they are often either time-consuming or merely quantify COD using test completion times. As it is known that markerless motion capture will be used during this project to gather richer information regarding agility, the metrics and associations with agility movements that can be visually distinguished are the most relevant. Therefore, force-related metrics will not be implemented. However, metrics related to joint angles, centre of mass, and ground contact time are assumed to be extractable from camera footage. Additionally, velocity (using the location over time) will also be included, as straight sprinting speed covers CODS in part (as can be seen in figure 1). Moreover, looking at the same figure, CODS is also partly explained by concentric strength and power. As these are required for fast acceleration and deceleration, these will also be derived from the velocity. The COD test that will be carried out is the arrowhead test because of its reliability in measuring CODS in a football context. It is also closely related to the AFCT test, which is already part of the training regime of both FC Twente and the FC Twente/Heracles Academy. Therefore, this test will also be part of the study design. However, using this test only has the purpose of allowing future expansion of the work presented in this thesis, as the test is yet to be validated.

As for the MMC methods, open-source pose detection software will be used to, besides it being cheap, make it easier for others to reproduce and verify the results of this thesis. It has become clear that each open-source option has its own unique advantages and disadvantages. As a system is required that only needs to be able to detect a single person, BlazePose might already suffice.

However, BlazePose has been shown to have a relatively low keypoint detection rate, and joint angle calculation using the software is expected to be less accurate. OpenPose and AlphaPose both seem better alternatives. Although AlphaPose is supposedly more accurate, OpenPose has a slight preference as it is assumed to handle occlusion better. This is crucial because the test that will be carried out has a set-up containing 2-meter high poles that participants must run around. There is no perspective the camera can be placed in that can prevent the occlusion of the participant. Although ViTPose is thought to be a good and simple alternative, OpenPose is thought to be at least similar in performance and ease of use. The final reason to use OpenPose is the knowledge of the system present within the Human Media Interaction department of the University of Twente and the fact that it should serve as a tool that is practical in use.

Chapter 3

Methods

The previous chapter concluded with a set of options that were chosen to be implemented as part of this thesis. In this chapter, the methods used to realize these choices are described, starting with the study design and participants. This is followed by the set-up used for capturing the motion of the participants taking the COD test. Then, the software that is written for data processing is detailed, followed by the feature extraction methods used. Finally, the data analysis of the features extracted from the data is addressed. The steps that are taken to get from recording the data to the analysis of the features are displayed in figure 4, which will be referred to in the related sections of this chapter.

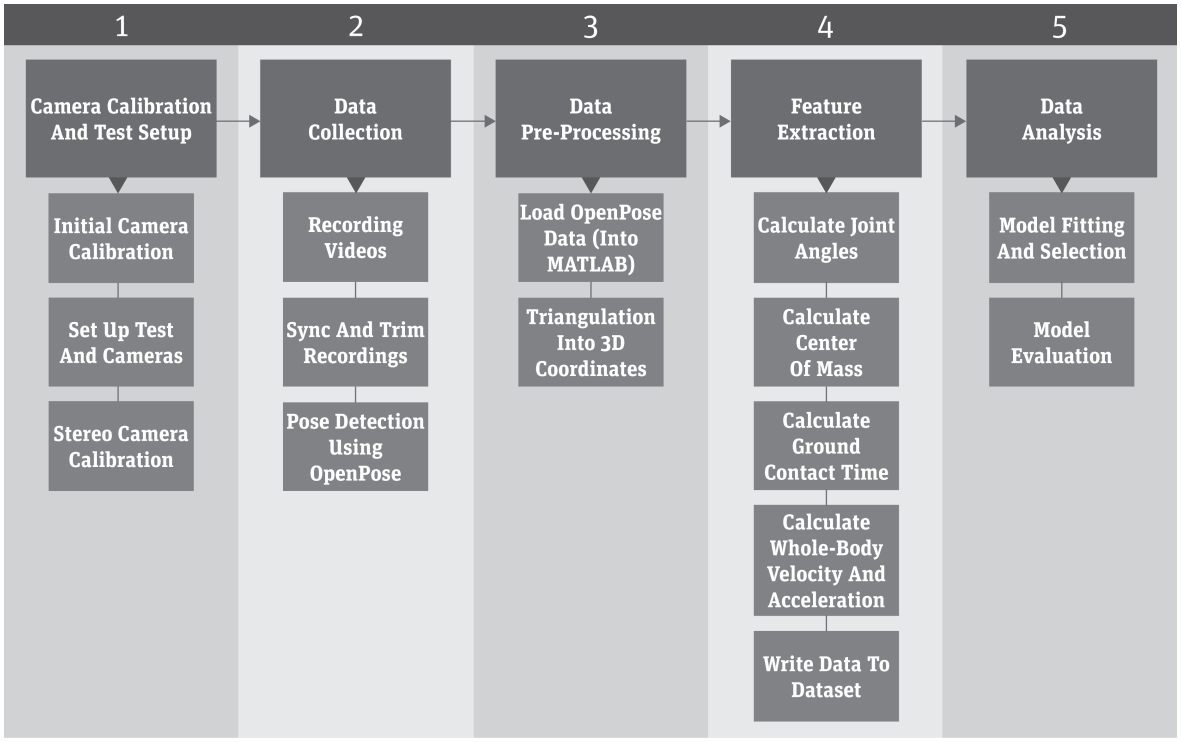


Figure 4: Steps in the development of the data pipeline that are followed within this thesis.

3.1 Study Design

To gather data about the rapid, multi-directional movements belonging to agility, participants performed the arrowhead test as described in section 2.2.1 due to its high validity and reliability as a benchmark for agility. Videos were simultaneously recorded from two perspectives so depth information could be derived. Additionally, to allow for future extension of the research, participants performed a second COD test, the "Agility FC Twente" test, again described in section 2.2.1. Furthermore, the participants' movements in both tests were recorded using a gold standard system for capturing full-body movement, allowing for future validation of the findings and developments

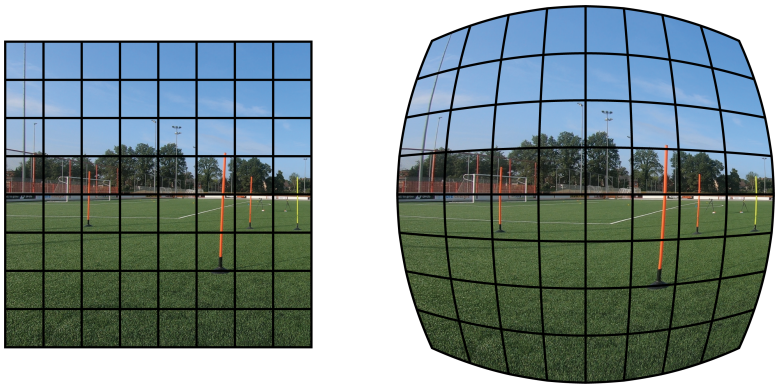
of this thesis. The system used for this is the MVN Awinda motion capture system based on inertial measurement units (IMUs) (Movella Inc., n.d.[b]). Movella Inc., the company behind this system, is a global leader in the digitization of movement (Movella Inc., n.d.[a]). "The reliability and validity of [the MVN Awinda sensors] for obtaining joint kinematics have been confirmed against gold-standard optoelectronic systems such as the Optotrak system and [...] the VICON system" (Cudejko et al., 2022). Prior to each test, the participants visually tapped an IMU-sensor on one of their thighs three times. This way, the exact moments in time of the taps can be found in the sensor data as large peaks in the sensor's corresponding segment acceleration, as well as visually in the video recordings. In work building upon this thesis, this can be used to synchronize all the data.

3.1.1 Participant Selection

A total of 50 participants were recruited by the FC Twente/Heracles Academy. The mean age of the participants was 17 (SD = 1.6). All participants were youth players of the academy, of which 13 were players of the Under-16 team, 22 were of the Under-18 team, and 15 were part of the Under-21 team. The participants were asked to sign a consent form drafted by the FC Twente/Heracles Academy in consultation with the University of Twente (see appendix A). The research has been approved by the ethics committee of the EEMCS faculty and is stored under reference number 230135. In the case of the under-aged participants, the parents were asked to sign the document. In a preliminary pilot test conducted beforehand, two adult participants performed the same test and signed the consent form available in appendix B.

3.2 Markerless Motion Capture Set-Up

The goal of using markerless motion capture is to capture the 3D movements of the test participants so that biomechanical information can be extracted from it. In order for this to work, a two-camera set-up is needed. Additionally, information regarding both cameras must be obtained to undo the lens distortion of the camera and to acquire depth information related to the captured images. This section is dedicated to describing this set-up, as well as detailing the methods used to extract the camera parameters through camera calibration. This section contains steps 1 and 2 of the data pipeline of figure 4.



(a) Undistorted image (b) Distorted image

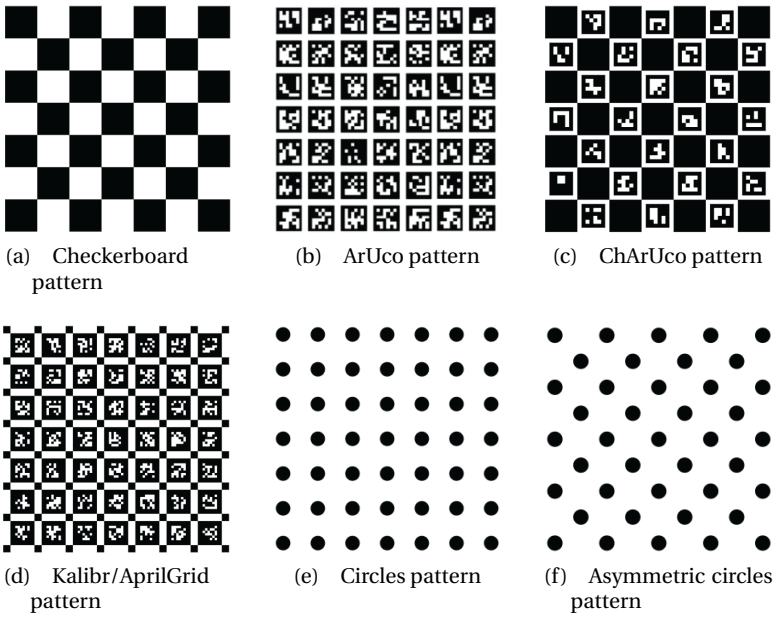
Figure 5: Example of how a camera lens distorts reality.

3.2.1 Background

As markerless motion capture does not use any on-body sensors or markers, its pose detection only relies on the camera images it is fed. However, most camera lenses distort reality to a certain extent, which, if not addressed, can result in strange results when the detected poses are analysed. An example of extreme lens distortion is shown in figure 5, which can be achieved through a fish-eye or a wide-angle lens. To correct such distortions, information about how the camera captures images must be extracted, such as the focal length, principal point, and image resolution. These are called camera intrinsics or intrinsic parameters. Some camera brands have this information

publicly available, but for some brands, they are more difficult to find. In those cases, these parameters can be estimated with tools like OpenCV (OpenCV, n.d.[a]) or MATLAB's Camera Calibrator App (The MathWorks Inc., n.d.). The information retrieved from camera calibration can, amongst others, be used to calculate the distance of an object to the camera.

To estimate the intrinsic parameters of the camera, images that include a calibration pattern are often used as input. Such a pattern is detected and the differences between the image pattern and the real-world pattern are compared. The calculations assume that the pattern that is used is completely flat, so it is important to use a surface as flat as possible to display the pattern. As can be seen in the example in figure 5b, the image is mostly distorted around the edges and the least in the centre. Therefore it is crucial that multiple images are used for the calibration, where the pattern is visible in all corners of the image. After estimating the camera parameters, they are used to project the calculated undistorted location of a point in the input image back onto the image. The difference in pixels between the detected point on the calibration pattern and the reprojected point is called the reprojection error. As its unit of measurement is pixels, the reprojection error scales along with the resolution of the input image. For individual video frames, the error tends to be much smaller than a single pixel (Eiríksson, 2022). However, when the image resolution is very high (e.g. 4K), an error of one or two pixels is also deemed acceptable.



(a) Checkerboard pattern (b) ArUco pattern (c) ChArUco pattern

(d) Kalibr/AprilGrid pattern (e) Circles pattern (f) Asymmetric circles pattern

Figure 6: Several calibration patterns that can be used for camera calibration.

Several calibration patterns exist that are all useful for camera calibration. An overview of these patterns is displayed in figure 6. A pattern that is straightforward and commonly used is the checkerboard pattern (figure 6a), where the distinctive corners of the squares are detected by the software. This is generally an accurate solution. Another pattern is shown in figure 6b. This is a matrix of ArUco markers (OpenCV, n.d.[b]), which are "synthetic square marker[s] composed [of] a wide black border and an inner binary matrix which determines its identifier" (OpenCV, n.d.[c]). Combining these two patterns results in the ChArUco pattern (figure 6c). The Kalibr/AprilGrid pattern (figure 6d) is fairly similar, as it also makes use of the ArUco markers. The detection algorithms of the ArUco, ChArUco, and Kalibr/AprilGrid patterns are slightly more complex, but because each marker can be uniquely identified, the pattern can still be useful even when only parts of it are visible (Wilm, n.d.). This is not possible with the checkerboard pattern, as the detected corners cannot be distinguished from each other. The remaining two patterns contain circles instead of squares and are displayed in figure 6e and figure 6f. In contrast to the checkerboard pattern, where only a few corner pixels in between the squares can be used to detect the pattern, all pixels surrounding a circle can be used to detect its location, which decreases the influence of image noise. Only the checkerboard, circles and asymmetric circles are currently supported by MATLAB's Calibrator App. OpenCV supports the same patterns with the addition of the ArUco and ChArUco patterns.

3.2.2 Camera Set-Up

As already established in section 2.5, the open-source software OpenPose is used for detecting human body poses. Although OpenPose offers whole-body 3D pose reconstruction and estimation, this functionality is currently only supported for (off-the-shelf) USB3 Flir depth cameras (Hidalgo et al., 2017; Teledyne FLIR LLC, n.d.). As this type of camera is less commonly used than other non-depth cameras, it is interesting to find out whether the same can be achieved using a normal, non-depth, off-the-shelf camera. For consistency, it was desired to at least use two cameras capable of filming in equal display resolutions and frame rates. Therefore, using multiple cameras of the same model was deemed practical. A camera model that was sufficiently available at the University of Twente was the GoPro Hero 7 Black (GoPro Inc., n.d.).

The camera configuration relative to the test set-up is displayed in figure 7. Here, the red and orange dots represent the 2-meter high poles that make up the route the participants run in the arrowhead test. This approximate route is displayed using the blue (left side) and green (right side) arrow-headed lines. The pink dotted lines represent boundaries behind which the test leader and any technicians and spectators should stand during the test. This made sure that there were as few people visible on screen as possible, which would make the data processing significantly easier, faster and more accurate. The boundaries were physically made part of the set-up using pylons, so it was clear to any bystanders where they were and were not allowed to stand.

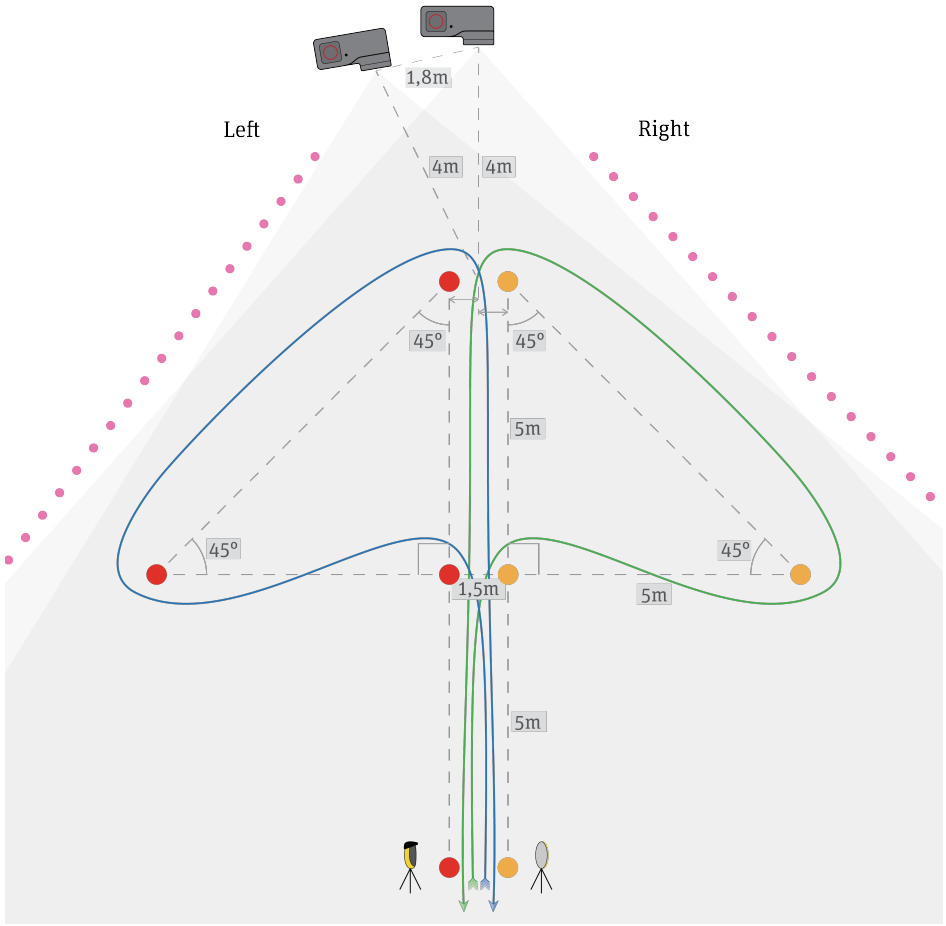


Figure 7: Camera and arrowhead test set-up.

Furthermore, it was found from preliminary testing that if the cameras were slightly angled down, the data points after triangulation into 3D-coordinates (further described in section 3.3.2) would appear further up the z-axis the further away the pose keypoints were from the camera. Illustrated in figure 8, the grey dotted line represents the horizontal viewing direction of the camera, as well as the height at which the body of the runner stays positioned in reality. Additionally, the red dotted line represents the viewing direction of the camera when it is angled down. In reality, the runner remains

at the same height, but from the camera's point of view, it moves up and away from its line of sight as he runs away from the camera. Hence, to acquire the correct keypoint height independent from the distance between the participant and the camera, the angle between the ground and the camera direction should be zero. Therefore, the camera's viewing directions were kept horizontal.

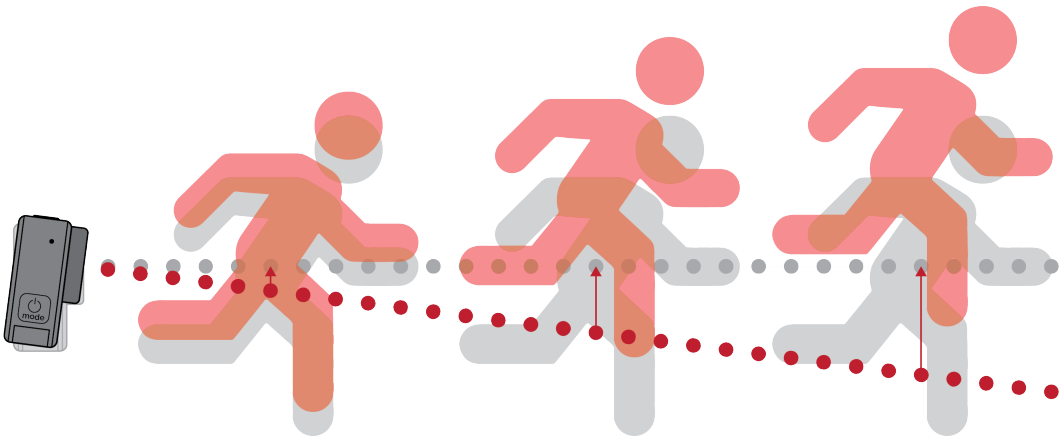


Figure 8: Effect of a downward camera angle on triangulated data.

3.2.3 Camera Calibration

To acquire the camera parameters, MATLAB's Stereo Calibrator App was used. This method was chosen because of its practical nature compared to the alternative methods described in section 3.2.1. Essentially, OpenCV can do the same, but it requires more manual input prone to human error. This is even more true for manually calculating all relevant camera parameters. However, using MATLAB's application comes with limited options for the calibration pattern. Of the options, the checkerboard calibration pattern was deemed the most practical because it could easily be created in the desired size using available software such as Adobe Illustrator. However, most patterns can also be generated using free tools such as Calib.io (calib.io, n.d.). In a preliminary test, a checkerboard of 9x8 (w x h) squares with sizes of 30 millimetres was used. However, two things became apparent from the test: (1) the calibration pattern was too small, and (2) the pattern could impossibly stay within the fields of view of all four cameras. The further away the calibration pattern was from the cameras, the more it was visible by the cameras, but the less accurately it could be detected. Hence, a larger calibration pattern was needed, and the cameras should not be placed too far apart. Therefore, a new checkerboard calibration pattern was created of 10x8 squares with sizes of 100 millimetres. The distance between cameras was set to 180 centimetres, as Zago et al. (2020) found that this results in the highest accuracy of markerless motion tracking using OpenPose. The authors reason that for the purpose of triangulating two 2d-coordinates into one 3d-coordinate, "the capture volume where the triangulated point can be placed" will decrease when increasing the between-camera distance. This, in turn, leads to a lower uncertainty in the triangulation process and, therefore, to a higher accuracy in the 3D reconstruction. Although Zago et al. (2020) researched this using straight gait tests, the principle still applies when applied to multi-directional change-of-direction tests, as the triangulation is performed separately for each key point in each frame of the videos.

The stereo images containing the checkerboard pattern were used to calibrate the cameras. As mentioned in section 3.2.1, the mean reprojection error should ideally be less than a pixel. However, the first few tests, in which both extrinsic and intrinsic parameters were estimated simultaneously, came back with errors much larger (>20 pixels). Therefore, a different approach was required, where the individual cameras were first calibrated separately to estimate their individual intrinsic parameters before doing the stereo calibration in the field to estimate the extrinsic parameters. This resulted in errors of about a pixel, slightly above the desired mean reprojection error. This is most likely caused by the high resolution of the images and the calibration pattern not being entirely flat. The same could be observed in the stereo calibration using the initial intrinsics of the two cameras. Here, the mean reprojection error (of 5 stereo calibrations in total) was between 0.7 and 1.5 pixels. However, judging by the locations and orientations of the cameras reconstructed by the MATLAB

calibrator, errors of the magnitude in the upper segment (i.e. >1.0 pixel) are still deemed acceptable for the purpose of this study.

3.2.4 Data Collection

The previous two sections were put into practice when all data used in this research was gathered on the third and fourth of July, 2023. The FC Twente/Heracles Academie organised two full days dedicated to the medical and physical screening of their players. Each player went through a nearly three-hour schedule, which included a medical screening, coordination tests, jump tests, sprint and COD tests, and ended with an endurance test (for Under-21 only). Each test has its own way of measuring the players' performances (e.g. time, height, or weight), but additional measurements were introduced to the sprint and agility tests. Player movement was measured using the IMU-based MVN Awinda motion capture system (Movella Inc., n.d.[b]) during the 30-meter sprint, the 5-0-5 agility test (505), the arrowhead test, and the agility test developed by FC Twente (AFCT). At the same time, the latter two agility tests were video-recorded using the set-up described in section 3.2.2, as these were of particular interest for this thesis. Unfortunately, the two days of testing were rainy days. This meant that, even though the camera lenses were wiped dry before each test, some video recordings were unusable due to raindrops on the lens. This caused the images to be too blurry for pose detection. If one video of a test is incorrectly recorded, then the full test becomes unusable. Additionally, some tests were not recorded due to technical issues. For example, the cameras sporadically crashed, which took too much time to wait for due to the tight schedule of the tests. Eventually, out of the 100 arrowhead tests recorded (two tests performed by 50 participants), comprising a total of 200 video recordings, 80 tests were correctly recorded.

Figure 9: Test and camera set-up on the days of data collection.

For each test, an instructor was present to explain how the test should be executed. Next to that, a second person was in charge of the camera set-up. Their tasks consisted of starting and stopping the recording for each trial, handling any errors that might arise throughout the day, and noting down the starting times of each trial. The latter was done to make it easier to match the two videos to each participant afterwards. In order to accelerate the process of syncing the videos afterwards, the *Multi Camera Control for GoPro* (Meyer, 2019) mobile application was used. The app allowed for the simultaneous starting and stopping of recordings on multiple GoPro cameras, which was tested for accuracy beforehand through a frame-by-frame analysis. The recordings were mostly correctly synced, yet some showed slight deviance of up to two frames, likely caused by small delays in the app-to-camera connection. However, this was deemed acceptable and relatively easy to correct. As a backup, a flashlight was used as a visual cue to synchronise videos manually afterwards. This was only used in case the connection with the app was lost, as this requires significantly more work in the processing of the videos. All video editing was done using Adobe Premiere Pro (Adobe, n.d.). As a final preparatory step before pose detection, the videos were trimmed so only the part where the test takes place remained.

3.2.5 Human Pose Detection using OpenPose

Once the videos were both equal in length and only included the segment of the test, they were ready for pose detection. The method chosen for this is OpenPose because it is expected to result in high detection rates, proper handling of occlusion, and because of the knowledge on the system already present in the university department (as stated in section 2.5).

There are numerous flags that can be used in OpenPose's command line interface to specify what should be detected and how this should be executed (Hidalgo et al., n.d.[a]). OpenPose is able to detect the face, hand and body of multiple people within the image frame. However, the face and the hand key points are not needed for the purpose of this study, as facial expressions and finger movement are too distantly related to agility. As for the key point detection of body poses, several model options exist. The default (i.e. when the *--model_pose* flag is not specified) is the BODY_25 model. However, an improvement of this model was made by Hidalgo et al. (2019), of which the first author is also the lead author of OpenPose. The differences between the two models can be seen in the left and middle images of figure 11. The improved BODY_25B model boasts improved "runtime performance while simultaneously improving slightly on the keypoint accuracy"(Hidalgo et al., 2019). Hence, this upgraded model was used.

The output is generated twofold. The *--write_video* flag is used to create a new video with a skeleton overlay of the poses detected in each frame. This can be used to check the accuracy of the pose detection based on face validity. However, the data that is required for further processing comes from the *--write_json* flag. Usage of this flag results in a separate JSON file for each frame of the input video. An example of such a file is displayed in figure 10. All empty elements are not used in this study. The data in the *pose_keypoints_2d* element is formatted as $x0, y0, c0, x1, y1, c1, ...$, where x is the x-coordinate of the keypoint, y the y-coordinate, and c the confidence score in the range $[0, 1]$.

```
{
  "version": 1.3,
  "people": [
    {
      "person_id": [ -1 ],
      "pose_keypoints_2d": [ 1443.07, 767.06, 1, 1443.06, 751.088, 1, ...],
      "face_keypoints_2d": [],
      "hand_left_keypoints_2d": [],
      "hand_right_keypoints_2d": [],
      "pose_keypoints_3d": [],
      "face_keypoints_3d": [],
      "hand_left_keypoints_3d": [],
      "hand_right_keypoints_3d": []
    }
  ]
}
```

Figure 10: Example JSON file

Furthermore, the *--number_people_max* flag is used to fix the number of people OpenPose should detect in each frame. Each test executed as part of this research contains only one person of interest, yet sometimes more people are visible inside the frame boundaries due to unforeseen circumstances. For example, in the preliminary tests, a coach was located inside the test area beside the participant who conducted the test. With the flag set to 1, OpenPose often switched between the participant and the coach when they were standing relatively close to each other. This is caused by the way OpenPose determines which person is of the most interest. When there are more people in view, the person with the highest "score" is kept, while all others are removed. OpenPose calculates this score based on the area the person covers within the image, as well as on the confidence values of individual joints and body parts (i.e. the certainty that a joint or body part is actually located at the resulting coordinates). Therefore, when the participant and coach were standing close to each other, the areas of the image frame their bodies covered were almost equal in size. It is highly likely that OpenPose was also similarly confident of individually detected joint and body part coordinates,

leading to pose detection of the participant in one frame and of the coach in the next. To minimise the probability that this effect reoccurred in future tests, people other than the to-be-tested participant were asked to stay out of view of the two cameras as much as possible (as was explained in section 3.2.2).

A somewhat related flag is the experimental *--tracking* flag. The flag, available although still in development, is meant for tracking people across frames, potentially resulting in higher accuracy pose detection. Preliminary tests showed slightly more stable detection of the person of interest when used together with the *--number_people_max* flag. Therefore, this flag was included in the definitive OpenPose settings used in this research. Lastly, the *--display* and *--render_pose* flags were found to be useful for increasing processing speed. The former ensures the video is not displayed when the flag is set to 0, while the latter stops the rendering of the skeleton overlay on top of each video frame that is displayed during execution. The latter does not influence the output of the *--write_video* flag.

3.3 Data Pre-Processing

With the data extracted from the videos using OpenPose, the data can be further processed to extract the 3D coordinates of the keypoints in real-world space for each frame of the video. Subsequently, this spatiotemporal information can be used to calculate whole-body velocity and acceleration, whole-body Centre of Mass (COM), ground contact time, joint angles, and asymmetries between the left and right leg. This section contains step 3 of the data pipeline of figure 4. It goes into detail about the process of going from two sets of raw x- and y-coordinates of OpenPose keypoints to usable 3D coordinates that form the basis of the biomechanical features calculated in the section 3.4.

3.3.1 Loading OpenPose Data

The JSON outputs of both cameras related to the same test participant and direction are consecutively loaded into MATLAB. Each JSON file is decoded into a structure array using the *jsondecode* (The Mathworks Inc., n.d.[c]) function, which transforms the JSON-formatted text into separately accessible nested fields. Of these fields, only the *pose_keypoints_2d* element is kept and split up into x and y variables (the confidence values are not used). The x- and y-variables from both perspectives all have the shape [*# of frames*] × [*# of keypoints*]. The first five keypoints are removed, as these are related to the gaze of the person, which is not needed for the purpose of this research. Additionally, OpenPose outputs data using zero-based indexing, while MATLAB uses one-based indexing. Therefore, the numbering of keypoints changes, influencing how they can be accessed in MATLAB.

In figure 11, the format of the BODY_25 (OpenPose default) and BODY_25B models can be seen, as well as the adjusted format of the BODY_25B model used in the remainder of this report (i.e. numbering adjusted to one-based indexing and with face keypoints removed). It is important to note some differences between the two models that must be taken into account when calculating biomechanical features later in the process. Especially for the COM calculation, which makes use of body segments (further explained in section 3.4.2), it is important to know the body landmarks associated with the keypoint positions of the model used. To start, keypoint 1 of BODY_25 is not located in the same place of the body as keypoint 17 of BODY_25B. The former is located in the midpoint between the shoulder joints, whereas the latter is located slightly higher at the upper vertebrae of the cervical spine. Furthermore, keypoint 18 of BODY_25B, located at the most cranial point of the body, is absent from the BODY_25 model.

3.3.2 Undistorting Points and Triangulation into 3D Coordinates

With the x- and y-coordinates accessible in separate variables, the data can be triangulated into 3D coordinates. However, OpenPose keypoints have been detected in the original distorted camera images and are, therefore, still distorted themselves. The *triangulate* function (The Mathworks Inc., n.d.[a]) that will be used to calculate 3D coordinates assumes that the points it receives as input are undistorted matching points in stereo images. Therefore, the 2D keypoint coordinates in each

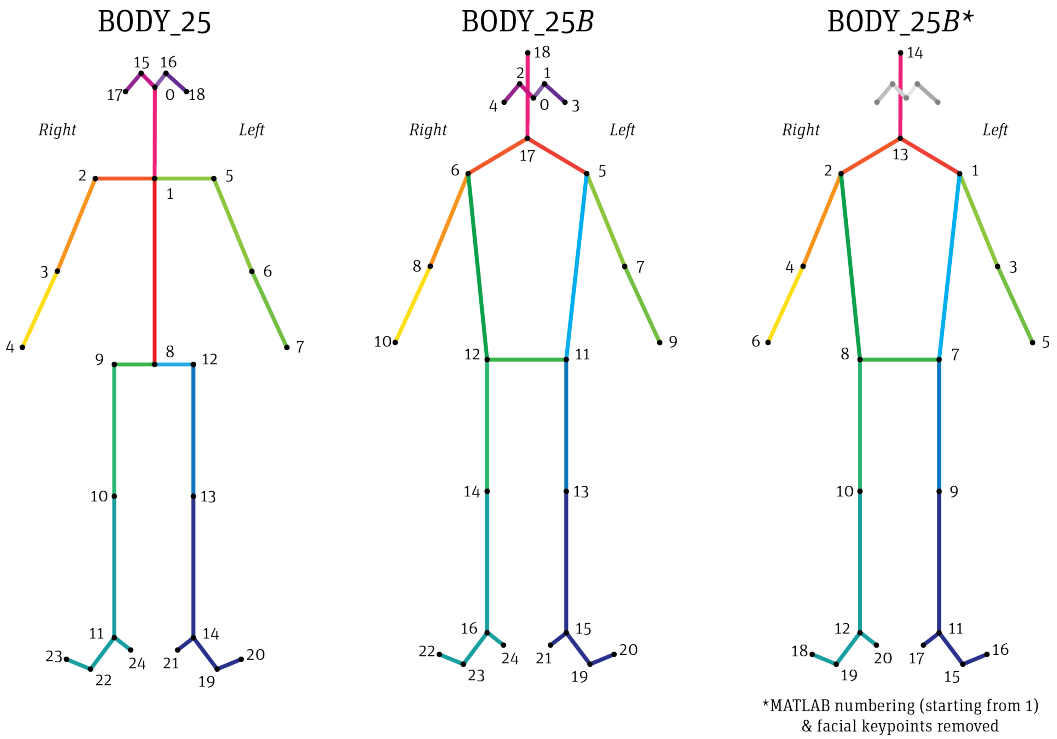


Figure 11: From left to right: BODY_25 output format, BODY_25B output format (used in this research), and BODY_25B output format with the facial keypoints removed and using MATLAB's one-based indexing.

camera perspective must first be undistorted using the individual camera parameters. The *undistortPoints* function (The Mathworks Inc., n.d.[b]) is used for this, which calculates the undistorted coordinates of the distorted points in one image at a time. As points need to be undistorted for all frames of both videos, the function is used in a loop over all video frames.

The *triangulate* function is used for each matching set of undistorted keypoints. For this, the x- and y-coordinates of a keypoint pair (i.e. the same keypoint but detected in two different perspectives) are used together with the set of camera parameters from the stereo calibration. The world coordinates are saved into separate variables: WX, WY, and WZ. All points are rotated -90 degrees around the X-axis to establish the correct orientation.

3.3.3 Scaling The Data

During the triangulation process, the data points are scaled so they can be expressed in millimetres in three-dimensional space. If this process were to be completely flawless, the real-world distances would be exactly equal to the triangulated distances. However, as error-free triangulation is highly difficult, a second step of scaling is required to achieve a more accurate display of distances and, therefore, of bodily proportions. To do this, de Leva (1996)'s adjustments to Zatsiorsky et al. (1990)'s segment inertia parameters are used. The scaling is done by adjusting all keypoints with a scaling factor based on the subject's real and virtual thigh lengths. The real thigh length is defined by

$$l_{thigh,real} = \frac{h_{real} * l_{thigh,dL}}{h_{male,dL}}, \tag{1}$$

where h_{real} is the participant's measured height in millimetres, and $l_{thigh,dL}$ and $h_{male,dL}$ are the male thigh length and male height from de Leva (1996)'s parameters in millimetres, respectively. The real thigh length was not measured directly from the participants as it was deemed inaccurate. De Leva (1996)'s segment lengths are measured between joint centres, which are difficult to pinpoint accurately from the naked eye. Moreover, measuring thigh lengths was not part of the screening process

of FC Twente, while measuring height was. The calculation is still an estimation – as would a measurement between the knee joint centre and hip joint centre be from the outside – but it is thought to be a better generalization of thigh lengths for the use of scaling. In the calculation, differences in body proportions between differently aged participants are not taken into account. The virtual thigh length is calculated step-wise. First, the virtual thigh lengths in each video frame are calculated by

$$d(i) = \sqrt{(h_i - k_i)^2}, \tag{2}$$

where $i = 1, \ldots, N$ in which N is the number of video frames, and h_i and k_i are the corresponding hip and knee joint locations, respectively. Then, the outliers are removed to clean the data of any inaccuracies. This is done using the *rmoutliers* (The Mathworks Inc., n.d.[d]) function, which removes data points that are more than three scaled median absolute deviations (MAD) from the median. The MAD is "a robust measure of how spread out a set of data is" (Glen, n.d.) and is defined as

$$MAD = median(|d_i - median(d)|), \tag{3}$$

where d_i refers to the virtual thigh length in video frame i. To ensure that the MAD is an unbiased estimator of the standard deviation of the data, the MAD is scaled by the scale factor k:

$$MAD_{scaled} = k \cdot MAD, \tag{4}$$

where k is a constant scale factor which, for normally distributed data, is defined as

$$k = 1/(\phi^{-1}(3/4)) \approx 1.4826. \tag{5}$$

Subsequently, the data is cleaned from the outliers following the formula

$$d_c = \{d_i \mid |d_i - median(d)| \leq 3 \cdot MAD_{scaled}\}, \tag{6}$$

where d_c is the cleaned data set of virtual thigh distances. Then, the mean is taken from the cleaned data

$$l_{thigh,virtual} = \overline{d_c}, \tag{7}$$

which is used in the final step in calculating the scaling factor

$$f_{scaling} = \frac{l_{thigh,real}}{l_{thigh,virtual}}. \tag{8}$$

Finally, the separate world coordinate variables WX, WY, and WZ are multiplied by the scaling factor. These three-dimensional coordinates will be used as the basis for all calculations described in section 3.4, which talks about feature extraction.

3.3.4 Removing Frames with Inaccurately Detected Keypoints

While checking the 3D reconstructions of participant movements, it quickly became clear that in approximately the first and last quarters of the test, the pose detection was rather poor. This was likely the cause of the participant being too distant from the cameras, which often caused the limbs to be indistinguishable from each other. This resulted in both legs being detected at the same position. Another problem surfaced at the locations where the participant is partly occluded by the poles marking the test route. Here, the foot of a pole is often detected as the foot of the participant, or the pole is even detected as a full leg. This is something to keep in mind when doing further research

on the topic at hand. Although OpenPose was expected to be able to handle occlusions sufficiently, they might just be a limitation of MMC.

There is no easy way to fix these parts of the data. However, the frames that are captured when the participant is closer to the camera can still be used. The occlusion problem was most present around the poles in the middle of the set-up (B_L and B_R in figure 2c). Additionally, most incorrectly detected keypoints corresponded to a distance beyond this point (i.e. further than approximately 9 meters from the camera). With these observations in mind, the data was trimmed down to only contain the part right after the participant has run around B_L or B_R, up until the point where they pass the B-line again. This is visualized in figure 12, where the orange part of the path is kept and used for feature extraction, while the grey part remains unused. In the following section, only the orange part is used for feature extraction.

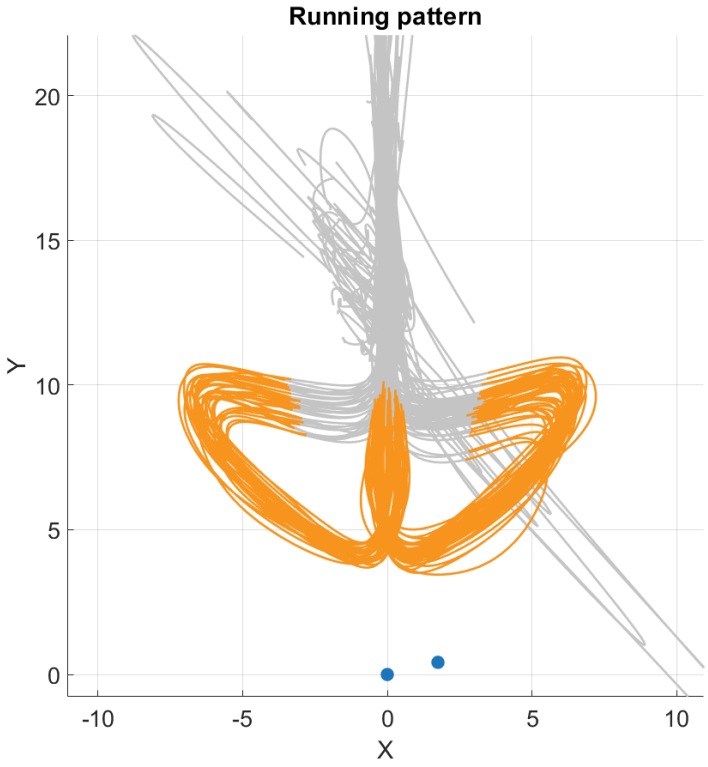


Figure 12: Running pattern based on the centre of mass position as seen from above. The orange part of the data is kept, and the grey part is removed. The blue dots represent the camera positions in the test set-up.

3.4 Feature Extraction

The previous sections laid the foundation required for the extraction of the biomechanical features mentioned in section 2.5. Building upon that basis of 3D keypoint coordinates over time, this section goes into depth about the calculations for obtaining ankle, knee, and hip angles, whole-body frame-by-frame centre-of-mass locations, heel strike and toe-off points in time, including a multitude of derivatives, and lastly, whole-body velocity and acceleration. With that, step 4 of the data pipeline of figure 4 is described.

3.4.1 Joint Angles

Joint angles are beneficial for calculating symmetry-related Key Performance Indicators (KPIs), such as the interlimb range-of-motion (ROM) difference. Joints that are deemed relevant for this way of quantifying agility are the hips, knees, and ankles. All other joints are not calculated as they are

Table 1: Joints with corresponding vectors between BODY_25B-numbered keypoints for calculating their joint angles.

Joint	Side	Keypoints Vector 1*	Keypoints Vector 2*
Hip	Left	11-5	11-13
	Right	12-6	12-14
Knee	Left	13-11	13-15
	Right	14-12	14-16
Ankle	Left	15-13	15-19
	Right	16-14	16-23

* Keypoint numbers are based on the original BODY_25B pose model, see middle image of figure 11.

deemed to be of lesser significance for said purpose. To calculate joint angles, the two connected vectors to the keypoint of interest are needed. Table 1 contains an overview of the vectors that are used to calculate left and right hip, knee, and ankle joint angles. Vectors mentioned in the table consist of the original OpenPose BODY_25B keypoints, of which the keypoint numbering can be seen in the middle image of figure 11. The shoulder and knee joints are used to calculate the hip angle. The hip and ankle joints are used for the knee angle. Lastly, the knee joint and forefoot are used for the ankle angle.

3.4.1.1 Disbalance

As it was found that lower limb asymmetry might be an indicator of agility (Bishop et al., 2021; Philipp et al., 2021), the disbalance between left and right joint angles is derived from the hip, knee, and ankle joint angles. This is done by saving the angle of each joint for each frame of the video. To enrich the information gathered about joint disbalance, the KPIs are divided into low-, mid- and high-range disbalance. Here, low-range is considered 0-60°, mid-range is 60-120°, and high-range is 120-180°. More than 180° could also occur, but it is very unlikely that this angle is present during an agility test. Therefore, the range of angles is limited to 180°. The KPIs are indicators of how balanced a joint is in the time span of the agility test. The disbalance is weighted against the magnitude of the angle range so that a difference between left and right of, for example, 4, weighs heavier in a total amount of 6 frames than it does in 60. For each joint and range of angles, the disbalance is calculated using the formula

$$d_{a,j} = 2 \cdot \frac{| N_{a,j,l} - \overline{N_{a,j}} |}{\sum N_{a,j}}, \tag{9}$$

where $d_{a,j}$ is the disbalance of joint j in angle range a, $N_{a,j,l}$ is the number of frames in which the left side of joint j is in angle range a, $\overline{N_{a,j}}$ is the mean of the number of frames in which both sides of joint j are in angle range a, and $\sum N_{a,j}$ is the total sum of the number of frames in which joint j is in angle range a. The multiplication by 2 in the formula functions to let the results range from 0 to 1. If, for example, for the duration of the test, the participant had their left knee in the low angle range for 16 frames ($N_{a,j,l}$) and their right knee for 20 frames, then $\overline{N_{a,j}} = 18$ and $\sum N_{a,j} = 36$, resulting in the disbalance

$$d_{a,j} = 2 \cdot \frac{| N_{a,j,l} - \overline{N_{a,j}} |}{\sum N_{a,j}} = 2 \cdot \frac{|16 - 18|}{36} = \frac{1}{9} \approx 0.111. \tag{10}$$

A value near 0 would mean a low disbalance, and a value near 1 a high disbalance. In the example, as $d_{a,j} \approx 0.111$, which is relatively close to 0, the disbalance is rather low.

3.4.1.2 Range of Motion

The measurement of the range of motion (ROM) plays an important role in rehabilitation (Davis et al., 2007) and improving sports performance (Lundgren et al., 2013). Therefore, it is interesting to find out whether there are correlations to be found between agility and hip, knee, and/or ankle ROM. The ROM is calculated as the difference between the largest and the smallest joint angle throughout the duration of the test. As asymmetry is expected to play an important role in agility, the ROM is calculated for the left and right joints separately, and the difference between the two is recorded as well.

3.4.2 Centre of Mass

The centre of mass of a set of point masses can be calculated by the weighted average of all the points (Bai et al., 2008). This follows the formula

$$COM = \frac{\sum m_i X_i}{\sum m_i}, \tag{11}$$

where m_i is the mass and X_i the location of point i. For the calculation of the COM of the body poses present in each video frame, m_i is taken from de Leva (1996). In their work, they present segment inertia parameters adjusted to joint centres as landmarks. This is different from the work from Zatsiorsky et al. (1990) they build upon, who make use of bony landmarks "markedly distant from the joint centers currently used by most researchers as reference points"(de Leva, 1996). The parameters consist of "relative body segment masses, center of mass positions, and radii of gyration"(de Leva, 1996). The former parameter is used as m_i, meaning the mass is used in the form of body weight percentages per segment i. The second parameter is used to calculate X_i following a formula adjusted from the work by Webering et al. (2021). The authors use OpenPose for measuring vertical jump height, meaning they only incorporate the vertical centre of mass positions. For the current research, the positions on all three axes are of relevance, resulting in an altered formula defined as

$$X_i = p_{start,i} + l_i^p (p_{end,i} - p_{start,i}), \tag{12}$$

where $p_{start,i}$ and $p_{end,i}$ are the coordinates of the start and end points of body segment i, and l_i^p is the "relative position of the segment's [centre of mass] along the principal axis of segment i"(Webering et al., 2021). The segment start and end points used, along with the associated segment masses and relative COM positions, are shown in table 2. For most of the segments, the regular endpoints as described by de Leva (1996) are used. However, alternative endpoints are used for the head and trunk, as the start and endpoints described for these segments were more fitting to the keypoints of the BODY_25B model. As for the head, the alternative segment starts at "the most cranial point of the head" (i.e. the highest point when standing upright) and ends at "the superior palpable point of the spine of the seventh cervical vertebra" (i.e. the top of the lowest vertebra of the neck, i.e. C7). These points are very close to keypoints 18 and 17, respectively, of the BODY_25B model. However, the regular head segment endpoints are not, as there is no keypoint close to the middle of the gonions of the jaw (i.e. the apex of the angle of the lower jaw), as seen from the frontal plane. Likewise, for the trunk, the segment includes either the top of the sternum, the C7 vertebra, or the midpoint between the shoulder joints. Because in the pose model, the only keypoint around the neck is located rather high up the cervical spine, the most cranial of the three possible endpoints is used, which is the C7 vertebra. The other endpoint is the middle of the hip (MIDH), which can be derived easily from the OpenPose data as the middle point between the hip keypoints. The hands are not used in the calculation because they are not part of the pose model and would require running OpenPose with the additional hand detector. Besides, the relative mass percentage of the hands with respect to the full body is negligibly small.

As a final step, the COM locations are filtered to remove obvious outliers and to reduce noise. Outliers are filtered using the *filloutliers* (The Mathworks Inc., n.d.[e]) function, which detects outliers in the same manner as the previously used *rmoutliers* function. Linear interpolation of neighbouring, non-outlier values was used to replace the outliers. Furthermore, the function was used to detect

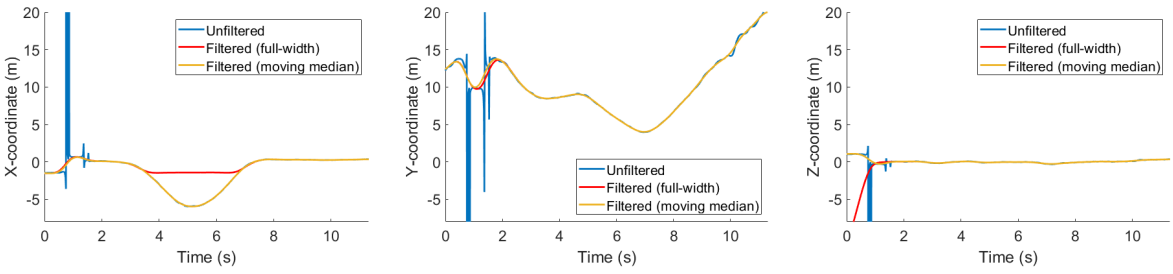
Table 2: OpenPose keypoints used as start and end points of body segments. Per segment, the body mass percentage (m_i) and the CoM's relative position along the segment's principal axis (l_i^p) are displayed.

	Segment			
Name	start	end	m_i(%)	l_i^p(%)
Head*	18	17	6.94	50.02
Trunk*	17	MIDH**	43.46	51.38
Upper arms	5,6	7,8	2.71	57.72
Forearms	7,8	9,10	1.62	45.74
Thighs	11,12	13,14	14.16	40.95
Shanks	13,14	15,16	4.33	44.59
Feet	21,24	19,23	1.37	44.15
Hands***	-	-	0.61	79.00

* Alternative endpoints are used for the head and trunk segments.
** Mid-hip, i.e. the centre point between hip keypoints 7 and 8.
*** The hands were not used in the COM calculation.

local outliers using the median of a moving window instead of the full width. Without the moving window, important information would be lost, as is shown in figure 13. In this example, the difference is most notable in the X-coordinates of the centre of mass, where the red line "cuts off" the trough in the original data.

Additionally, a second-order Butterworth filter(Parks et al., 1987) was applied to the data using the video's framerate (i.e. 60 fps) as the sampling rate. A Butterworth filter is often used for filtering out noise in kinematic data (Beckett et al., 2017; Nakano et al., 2020; Sakurai et al., 2021; Washabaugh et al., 2022), which makes it highly applicable to the noise likely caused by the slight differences in OpenPose's keypoint detection in between frames. The cutoff frequency of $1Hz$ and the order were chosen following an iterative process to achieve the desired filter effect. Figure 14 contains an example of where the centre of mass is located (after filtering) with respect to the 3d-reconstructed pose, accompanied by the corresponding frame.



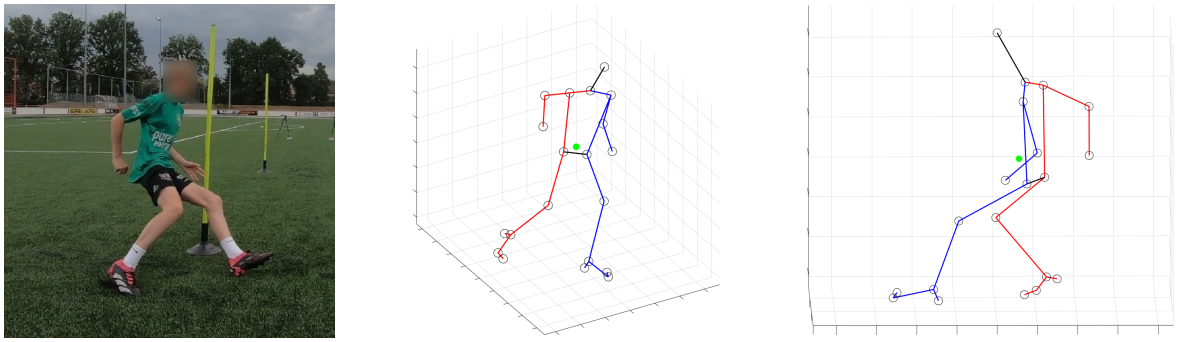
(a) CoM X-coordinates (b) CoM Y-coordinates (c) CoM Z-coordinates

Figure 13: Filtered and unfiltered centre of mass XYZ-coordinates over time.

3.4.2.1 Take-Off Distance

A relatively large distance between the COM and the foot placement of the trail leg has been associated with improved CODS (Hewit et al., 2012; Welch et al., 2019). Therefore, this take-off distance can be calculated as a derivative of the centre of mass and the pose keypoints. In a technical analysis of a 180° CoD task, Hewit et al. (2012) consistently observed a large take-off distance in superior performances. The authors reason that a large take-off distance results in a large step length, and as long as the step frequency is at least maintained, this, in turn, results in increased velocity. The take-off distance is calculated for each video frame in which the toes of the trail leg leave the ground. How these toe-off frames are determined is explained in further detail in section 3.4.3. The distance

between the centre of mass location, as calculated in the current section, and the keypoints corresponding to the toes (keypoint 20 for left and 22 for right) make up the take-off distance.



(a) Still from COD test video

(b) 3D reconstructed pose (first perspective)

(c) 3D reconstructed pose (second perspective)

Figure 14: 3D reconstructed pose with calculated centre of mass locations indicated by the green dot.

3.4.3 Ground Contact Time

A shorter ground contact time has been linked to greater CODS (Condello et al., 2016; Dos'Santos et al., 2017; Welch et al., 2019). Using accelerometers, Purcell et al. (2005) has proven it is feasible to calculate the ground contact time during running. Their approach provided "very close estimates of actual [contact time] during running", of which the best estimates were obtained during the highest measured accelerations. OpenPose keypoints provide rather similar information to accelerometer data. While accelerometers measure acceleration, in m/s^2, OpenPose data can be processed to acquire positional data over time (i.e. velocity), of which the derivative is acceleration. It should be noted that when calculating derivatives based on a series of data points, any inaccuracies present in the data might exponentially increase. For example, when a keypoint is detected in the wrong location in one frame, the calculated speed at that point in time will be enormous. The acceleration there will be even greater. Nevertheless, the idea is implemented for human gait analysis by Stenum et al. (2021). However, they calculate the events only from a lateral view, meaning that the anterior-posterior distances they use in their calculations can be directly derived from the X-coordinates of the ankle keypoints. In three-dimensional space where the participant does not walk in a straight line, this is more difficult to accomplish. Therefore, the way of determining event times in two-dimensional, one-directional gait cycles is altered to fit the three-dimensional, multi-directional change-of-direction tests of this research.

To achieve the desired result, the vertical positions of each keypoint are used to calculate the velocity in the vertical direction. Figure 15 contains plots of the vertical velocity of the heel and toe keypoints over time. The velocity of the heel should be (very close to) zero when the heels are first touching the ground, but also when the feet are at the highest point while raising up the legs. Therefore, for all negative peaks in vertical heel velocity, the height of the keypoint is used to determine which peak belongs to an actual heel strike and which peak corresponds to a leg raise. The peaks of which both neighbouring peaks corresponded with a greater vertical height were determined to be the heel strikes (i.e. the blue dots in figure 15). To calculate the corresponding toe-off events, the positive peaks in the vertical velocity of the toe keypoint in between the current heel strike and the subsequent leg raise moment are assessed. If only one peak is measured, then the frame corresponding to that peak is determined to be the toe-off event. However, if there is more than one peak (likely caused by faulty data), the highest peak is assumed to be the right frame for the toe-off event. Likewise, if there is no peak, the first vertical toe velocity peak found after the heel strike is used as the toe-off event. The toe-offs are displayed as orange dots in figure 15. Throughout the development of this method, the corresponding video frames were visually checked for face validity. It was found that it sometimes occurs that OpenPose detects the left leg as the right leg or vice versa. When this happens only in one frame, the pre-processing steps handle the problem quite well by replacing the faulty keypoint with interpolated keypoint coordinates. However, when the legs are switched

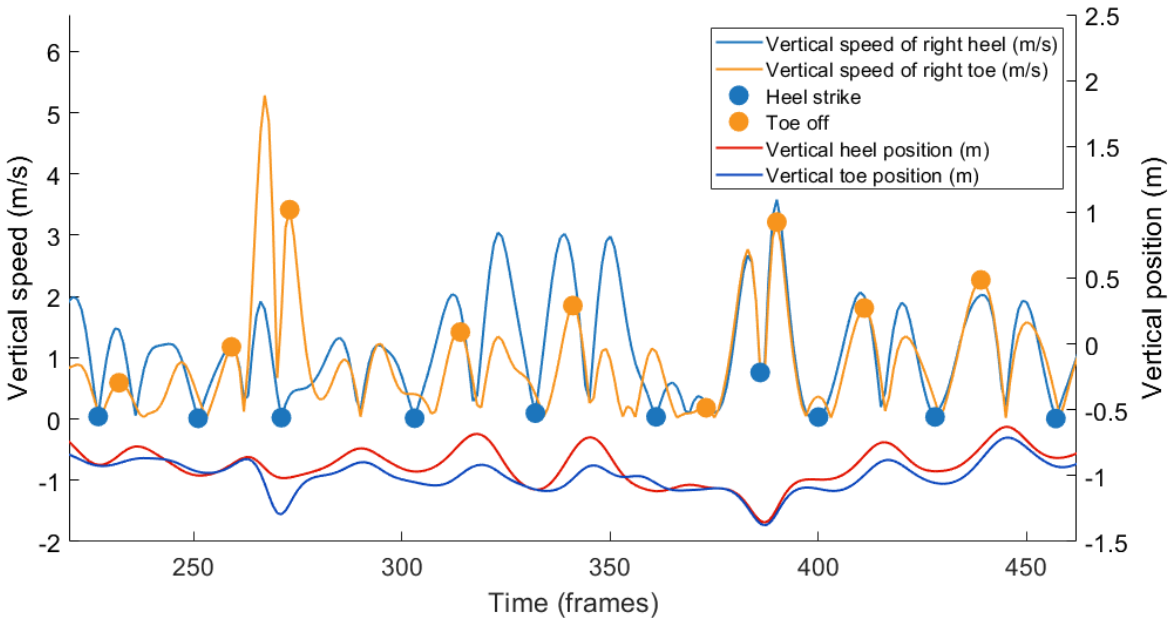


Figure 15: Participant 22 during a left-turned arrowhead test as an example of how heel strike and toe-off events are determined.

in several frames in a row, the problem is less accurately addressed. Although these are mostly intercepted by removing outliers that evidently display incorrect values (i.e. where the value is more than three scaled MAD from the median), there is still room for improvement here. However, this is primarily an issue of pose detection that might, therefore, be best addressed by OpenPose.

The events are used as the basis for a number of spatiotemporal parameters inspired by Stenum et al. (2021), which are described in the following subsections. For a visualization of the temporal parameters, see figure 16. Here, *lhs* is the left heel strike, *rhs* is the right heel strike, *lto* is the left toe-off, and *rto* is the right toe-off.

3.4.3.1 Stance Time

The stance time is the duration in which the foot is in contact with the ground (ergo, the actual ground contact time). As described by Stenum et al. (2021), it is calculated as the "duration in seconds between heel-strike and toe-off of the same leg". For the left and right stance time, this is the duration between *lhs* and *lto*, and between *rhs* and *rto*, respectively (see figure 16). The KPIs that are derived from this data are part bilateral and part unilateral. That is, the mean, minimum and maximum bilateral stance times are calculated based on all stance times, while the unilateral mean, minimum, maximum and total stance times are calculated for left and right separately.

3.4.3.2 Step Time

The step time is calculated as the "duration in seconds between consecutive bilateral heel strikes" (Stenum et al., 2021). For the left step time, it is the duration between *lhs* and the previous *rhs*. For the right, it is the duration between *rhs* and the previous *lhs*. In a likewise manner as for the stance time, the step time KPIs exist of bilateral and unilateral mean, minimum and maximum values. While the total time was included for the stance time, it was not calculated for the step time as this did not seem to be a very meaningful parameter.

3.4.3.3 Swing Time

The swing time is the time a foot is not touching the ground. It is calculated as the "duration in seconds between toe-off and heel-strike of the same leg" (Stenum et al., 2021). For the left swing time, it is the duration between *lhs* and the previous *lto*, and for the right between *rhs* and the previous *rto*. The swing time KPIs were the same as those of the step time.

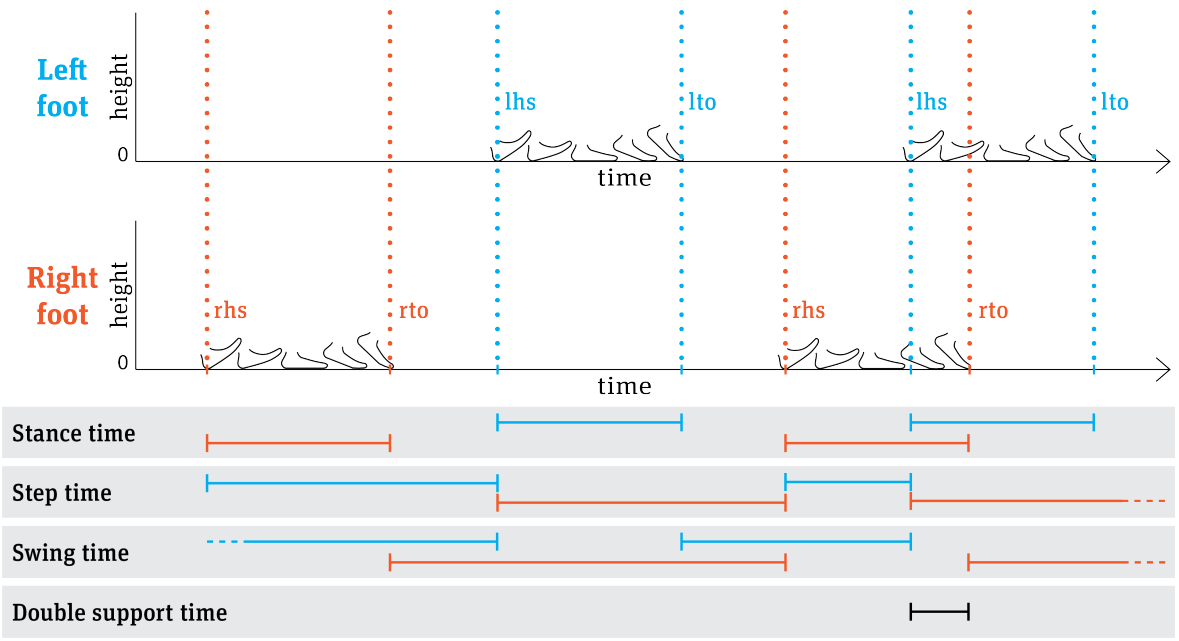


Figure 16: Left and right heel strike and toe-off events (*lhs, rhs, lto*, and *rto*, respectively) with indications of how temporal parameters are derived.

3.4.3.4 Double Support Time

The double support time is the time in which both feet are in contact with the ground. It is calculated as the "duration in seconds between heel-strike of one leg and toe-off of the contralateral [i.e. the other] leg"(Stenum et al., 2021). This is calculated from left to right and from right to left. The former is the duration between *lto* and the previous *rhs*, and the latter is the duration between *rto* and the previous *lhs*. For this parameter, looking at figure 16 might be especially insightful. While calculating this parameter, it often resulted in negative values. This means that the toe-off of one foot happens before the heel strike of the other foot. Therefore, a negative double support time, in fact, can be regarded as "flight time", where no feet touch the ground. This is not surprising, as both feet touching the ground simultaneously is less frequent, or even absent, during sprint tests or agility/COD tests. While the parameter is relevant for gait analysis, it might also be interesting to investigate for the purpose of this thesis. As for the KPIs, the total double support time was calculated. This was, in turn, divided by the completion time of the corresponding arrowhead test to arrive at an additional double support percentage KPI.

3.4.3.5 Step Length

The step length is the distance between a heel strike of one leg and the following heel strike of the other leg. It is calculated as the "anterior-posterior distance in meters between left and right [...] ankle keypoints [...] at heel-strike"(Stenum et al., 2021). However, for the multidirectional COD tests, the anterior-posterior distance is replaced by the Euclidean distance. The step length KPIs are equivalent to the step and swing time KPIs.

3.4.3.6 Running Gait Speed

Velocity can be calculated in other ways than by using the heel strike and toe-off events (see section 3.4.4), but the running gait speed is calculated with just that. As speed is often expressed in meters per second, and the step lengths and step times are known, the speed can be calculated by dividing the mean step length by the mean step time. This is the only running gait speed KPI that is used in the modelling phase.

3.4.4 Velocity and Acceleration

CODS is a measure that is partly covered by a person's straight sprinting speed and concentric strength and power present in accelerations and decelerations (see figure 1). Therefore, the velocity and acceleration during the COD test are interesting measures to calculate and investigate. For this thesis, the two measures are based on the location of their centre of mass, calculated according to section 3.4.2. For each frame, the whole-body velocity is calculated using the formula

$$v = \frac{\sqrt{d_x^2 + d_y^2 + d_z^2}}{t}, \tag{13}$$

where d_x, d_y and d_z are the displacements in x-, y-, and z-direction, and t is the time in seconds. The Euclidean distance between the centre of mass position in the previous frame and its position in the current frame is divided by time interval t. As the calculation is done for each frame of the video with a framerate of 60 frames per second, $t = 1/60$. The resulting value is in mm/s, which, after being divided by 1000, results in a velocity in m/s. The acceleration of a participant is computed by calculating the gradient of the velocity data series over the same time interval t using MATLAB's *gradient* (The Mathworks Inc., n.d.[f]) function.

To validate whether the computed velocity and acceleration are within the boundaries of what is physically possible, the data was plotted and compared to reference values based on Postma et al. (2022). The work proposes a model on (sprint) running kinematics, including action boundaries regarding velocity and acceleration. The maximum velocity and acceleration presented through the model are plotted as reference lines in the velocity and acceleration graphs over time in figure 17, where maximum velocity $v_{max} = 9m/s$, maximum acceleration $a_{max} = 6m/s^2$, and the maximum deceleration $a_{min} = -8m/s^2$. The latter is not presented in the work of Postma et al. (2022) but was retrieved from the website of StatSports (Daykin, n.d.). The grey lines show the full velocity and acceleration data, while the coloured lines show the part of the data that is actually used after trimming the data as described in section 3.3.4.

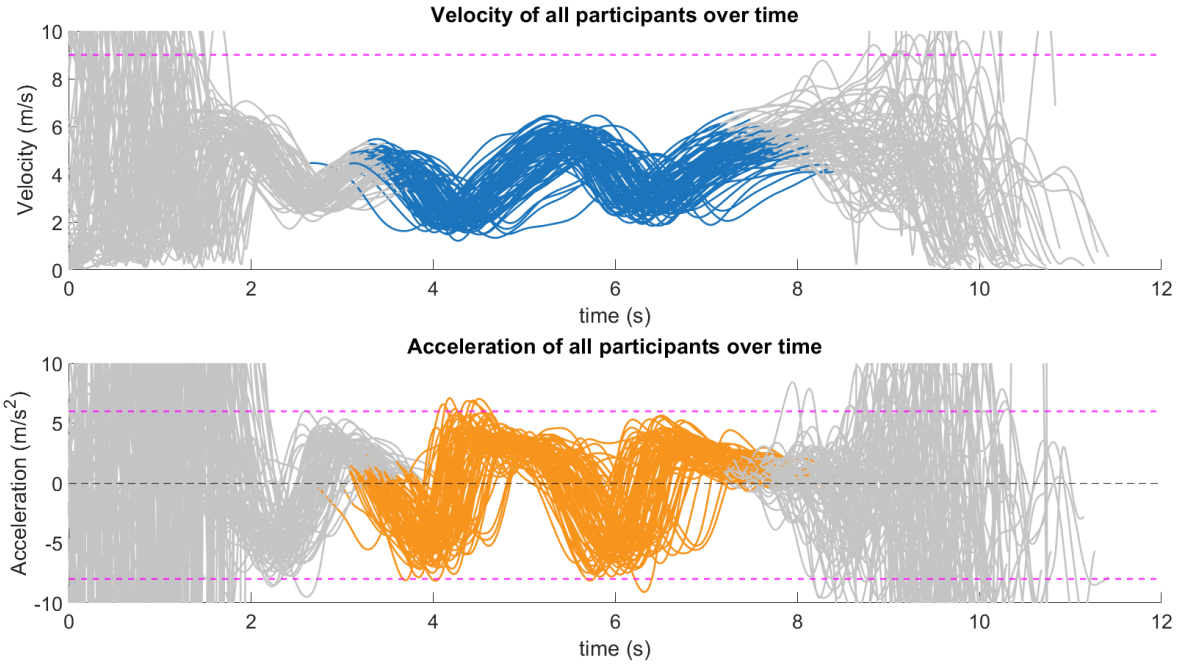


Figure 17: Whole-body velocity and acceleration of participants during their COD tests with reference values for validation.

3.4.5 Resulting Dataset

All of the KPIs are exported from MATLAB to an Excel spreadsheet automatically when running the complete script available on Github (Peetsma, 2023). The dataset comprises a total of 72 OpenPose-based variables. Furthermore, the OpenPose data was combined with another three anthropometric variables, one demographic variable and 20 variables from other tests that were conducted on July 3rd and 4th. The full set of variables, including their unit of measurement, can be found in appendix C.

3.5 Data Analysis Using GLMM

Following the data pipeline of figure 4, after the features are extracted from the video recordings, the data can be analyzed. This is done by modelling the parameters in the dataset to predict a completion time as close to the observed value. However, some background information is required first to understand the methods used. The section concludes by critically checking assumptions associated with mixed models. The remaining part of the model evaluation (i.e. the second part of step 5) will be addressed in the next chapter.

3.5.1 Background

Generalized Linear Models (LM) can be used to describe the relationship of a certain outcome variable with one or more parameters. A linear model "assumes that there is approximately a linear relationship between X and Y" (James et al., 2013). When Y is observed, and there are multiple X parameters (called a fixed effect when used in an LM) that might have a significant relationship with Y, then LMs can serve as a way to predict this relationship. For an LM approach to result in meaningful outcomes, all observations must be independent (Winter, 2013). When this is not the case, a more extensive method is required: Generalized Linear Mixed-Effect Models (GLMM). As opposed to an LM approach, Generalized Linear Mixed-Effect Modelling (GLMM) includes not only fixed effects but also random effects: an uncontrollable factor that stands for all that affects the outcome variable (also called the dependent variable) that is not captured in any fixed effect. GLMMs can be described by the formula defining their linear relationship (Bates et al., 2015):

$$Y \approx \beta_0 + \sum_{i=1}^{n}(\beta_i X_i) + \sum_{j=1}^{n}(Z_j) + \epsilon_j, \tag{14}$$

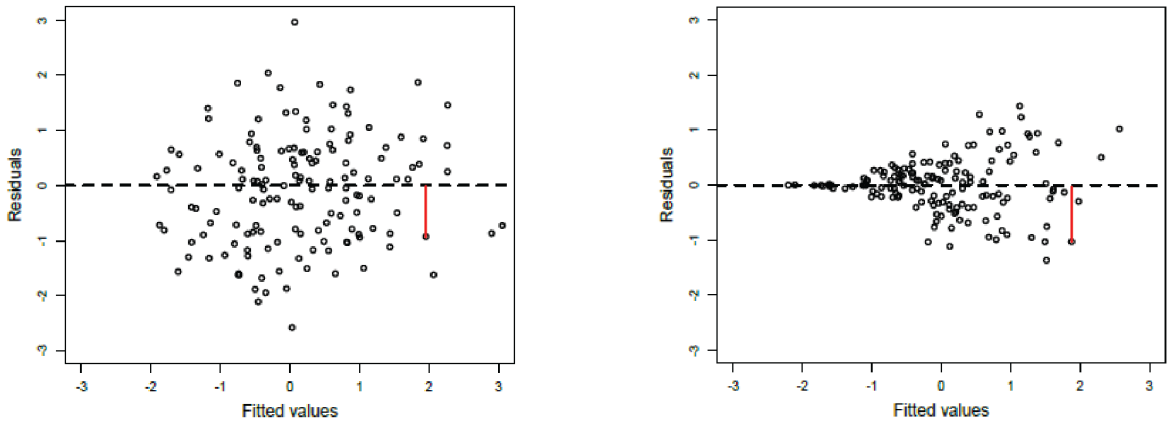
where Y is the outcome variable, β_0 is the *intercept*, β_i is the *slope*, X_i is a fixed effect, also called an independent, explanatory, or predictor variable, Z_j is the random effect for variable j, and ϵ is an error term (assumed to be independent of X (James et al., 2013)). The error term can also be interpreted as the unexplained or residual variability. The intercept and the slope are together referred to as the *model coefficients*.

3.5.1.1 Assumptions

There are a number of conditions that have to be satisfied to make sure that the linear model will be meaningful (Winter, 2013). Each of them is described below.

Linearity First, the model must be the result of a linear combination of the elements used as fixed and random effects. This can be verified using a residual plot. In figure 18, each observed value (i.e. the horizontal line) and its corresponding value predicted by the model (i.e. the points) can be plotted. The differences between them are the residuals, shown as the red lines. If a non-linear or curved pattern can be observed in the residual plot, the data would violate the linearity assumption. If not, like in figure 18a, then the assumption is respected.

Absence of Collinearity The second assumption is that the fixed effects used in a model are not collinear (i.e. correlated with each other). When multiple fixed effects are correlated with each other, the phenomenon is called multicollinearity. Should two collinear variables be used in the



(a) Data with homoscedastic variability (b) Data with heteroscedastic variability

Figure 18: Examples from Winter (2013) of residual plots, where the red line indicates a residual. (b) shows larger residuals for higher fitted values, whereas in (a) the residuals are approximately equal across the range of fitted values.

same model together, then the model has no proper way of determining to which of the variables the variance the model explains should be assigned to. Therefore, the model will not be easily interpretable. To find out which combinations of variables should be avoided, a correlation matrix of all available variables can be created. Variables with a collinearity coefficient $r > 0.7$ are considered collinear and should not be used in the same model. As a final test, the variance inflation factor (VIF) can be computed, which measures "how much the variance of a regression coefficient is inflated due to multicollinearity in the model"(Kassambara, 2019). The VIF can have a value down to 1, indicating there is absolutely no collinearity between the fixed effect and the other coefficients. An often-used rule-of-thumb is to regard any VIF higher than 5 (or even 10) as an indication of a problematic amount of collinearity.

Absence of Heteroscedasticity Thirdly, the data should be homoscedastic, meaning that "the variability of [the] data should be approximately equal across the range of [...] predicted values." (Winter, 2013). Figure 18 shows an example of how data with homoscedastic variability looks (desired) in comparison with heteroscedastic variability (undesired). The situation shown in figure 18b essentially means that the model predicts less accurately for higher values of the outcome variable, which is unwanted.

No Influential Data Points Lastly, the data should have no influential data points. A data point or observation is influential when it, "either individually or together with several other observations, has a demonstrably larger impact on the calculated values of various estimates [...] than is the case for most of the other observations" (Belsley et al., 1980). Influential observations can "drastically change the interpretation of [the] results, and [...] it can lead to unstable results" (Winter, 2013). To see whether a data point is influential, the DFBETAS metric can be evaluated. DFBETAS is "the change in parameter estimates after deleting the ith observation" (Li et al., 2011). In other words, the metric quantifies the difference between a model with a certain observation and the same model without that observation. According to Winter (2013), any value that changes the sign of the regression coefficient (i.e. the slope) requires particular attention. However, Belsley et al. (1980) mention a "size-adjusted cutoff" that can be used of $2/\sqrt{n}$, where n is the number of observations. Additionally, Winter (2013) mentions that any DFBETAS value that changes the slope of the coefficient is definitely an influential data point.

Normality of Residuals The assumption that the residuals are normally distributed is deliberately excluded. The assumption is that the residual data should be symmetrical around the mean residual, and data points close to the mean should be more frequent in occurrence than data more distant from the mean. However, linear models are relatively robust against normality violations, and the importance of this diagnostic is even debated (Winter, 2013). Therefore, this assumption will not be tested.

3.5.1.2 Evaluation Methods

Several methods can be used to assess the model fit. These can be divided into numerical and visual evaluations and are described in the subsection below.

Coefficient of Determination A model can be evaluated by finding how well the model is able to replicate the observed value. In statistics, this is expressed as the coefficient of determination, or R^2 (James et al., 2013), and can be calculated in numerous ways. After an assessment of the problems associated with the most common definitions, Nakagawa et al. (2013) proposed a method for calculating two types of R^2: marginal R^2, which represents the variance explained by the fixed effects, and the conditional R^2, representing the variance explained by the whole model, including fixed and random effects. When the conditional R^2 is close to 1, it means that a large proportion of the variability in the observed variable can be explained by the model. When the marginal R^2 is close to the conditional R^2, a large proportion of the explained variance comes from the fixed effects.

A study similar to this thesis was conducted by Rosenblum et al. (2023), who investigated the effect of arm restriction on dynamic stability and upper body responses to lateral loss of balance during walking. They report a marginal R^2 of 0.252 and a conditional R^2 of 0.611, which they interpret as a substantial association between their fixed effects and their outcome variable. Another study, by Sanchez et al. (2021), investigated several biomechanical variables in relation to post-stroke step length asymmetry (SLA). They report a marginal R^2 of 0.59 and a conditional R^2 of 0.84 for a model predicting SLA for participants who took longer steps with their paretic leg. For another model for participants walking with shorter paretic steps, the marginal R^2 was 0.19, and the conditional R^2 was 0.77. This is interpreted as an indication that there are additional differences unique to the subject with shorter paretic steps that are not accounted for by the variables included. Lastly, a study by Malik (2022) reported marginal R^2's of several models and related the values to the extent to which the fixed effects modify group effects. In other words, they discuss how much the fixed effects are able to explain the variance in the dependent variable for the whole population. According to the author, marginal R^2 values of 0.141 and 0.434 "may modify group effects", whereas values of 0.598, 0.766, and 0.808 were said to show an overall group effect. The results and interpretations of these studies can be used to determine the implications of the model to predict agility.

Residual Standard Error (RSE) The *residual standard error* (RSE), or *model sigma*, is a measure of the error of prediction defined by "the average amount that the response will deviate from the true regression line" (James et al., 2013). To calculate the RSE, the *residual sum of squares* is required, which, in itself, is already a metric to understand how well a model fits a dataset. It is defined by

$$RSS = \sum (e_i)^2, \tag{15}$$

where e_i is the ith residual. Using the RSS, the RSE can be calculated according to the following formula as described by James et al. (2013):

$$RSE = \sqrt{\frac{RSS}{(n-c)}}, \tag{16}$$

where n is the number of observations, and c is the number of model coefficients excluding the intercept (thus, c only includes the fixed effects included in the model). Divided by the mean value of the observed variable gives the prediction error rate (Kassambara, 2019). The RSE has the same

unit of measurement as the outcome variable, and the error rate is a percentage. A more accurate model will have a lower RSE value and a lower error rate.

Akaike Information Criterion (AIC) As an extension of the maximum likelihood principle, Akaike (1992) introduced the Akaike Information Criterion (AIC). Although it was the first model selection criterion to gain widespread attention, it is still one of the most used model selection tools to date (Cavanaugh et al., 2019). As a model can better adapt to slight differences in a dataset once the model gets larger and more complex, a complexity penalization is introduced to the statistic. Therefore, the AIC is a measure of both model quality and complexity. However, "it is not the absolute size of the AIC value, it is the relative values over the set of models considered, and particularly the differences between AIC values, that are important" (Burnham et al., 2004). A common rule of thumb is to regard a model as a viable candidate when its AIC value is reduced by a minimum of 2 units compared to the model with the previously lowest AIC (Wieling, 2018). This will ensure that the simplest model is selected that simultaneously best explains the variance of the dependent variable.

Likelihood Ratio Test Two different models can be compared "to determine if one is a better fit to the data than the other" using a likelihood ratio test (LRT) (Luke, 2017). An LRT compares the likelihoods of two models. This can be useful in case it is not sure whether to add or keep a parameter in a model or whether to remove it. When one model contains the parameter of interest and another model does not, the "fixed effect is significant if the difference between the likelihood of these two models is significant" (Winter, 2013). The model without the fixed effect in question is called the null model, and the one including it is called the full model.

3.5.2 Approach

As mentioned in section 3.5.1, Linear Models (LM) require all observations to be independent. In the case of this study, however, these observations cannot be assumed to be independent of one another as participants performed the COD test twice. Therefore, to predict how agile a participant is based on one or more predictor variables, Generalized Linear Mixed-Effect Models (GLMM) will be used.

To account for nested dependencies in the data resulting from the fact that every participant performed the test multiple times ($n = 2$), a random effect is introduced for the participants (PID). This effectively allows the model to find a different intercept for each participant and lets the model know that multiple observations of the same PID should be regarded as dependent. This builds on the assumption that each participant has a different "baseline" completion time. The fact that a participant is naturally fast affects both their (left and right) tests. The observed variable Y is the completion time of the COD test that corresponds with the OpenPose data analysis. The KPIs described in the previous sections are used as fixed effects X. RStudio (R Core Team, 2023) is used with the addition of the *lme4* package (Bates et al., 2015) to allow for GLMM fitting.

Before any fixed effects are introduced into a model, a 'base' or null model is created that includes purely the random effect for participants. Then, a matrix of collinearity coefficients was calculated for all variables in the dataset. Pairs of variables for which the coefficient r was above the threshold of 0.7 were regarded as collinear. Using a step-wise forward selection method, all variables derived from OpenPose were added to the intercept-only model one by one. A variable was kept in the model only when the model AIC reduced a minimum of 2 units compared to the model with the previously lowest AIC (i.e. $\Delta AIC < -2$) and when all the model's fixed effects were significantly correlated with the test completion time ($p < 0.05$). The models that met these criteria were continued with, and variables were once more added to them to check whether they would improve the model's explained variance. This process was continued until no better model could be found anymore. Next, the available anthropometric and demographic variables were added to the best models found. Of these, only the age turned out to be significant. The diagnostics of the remaining best-fitting models are displayed in table 3. The marginal R^2 of the two models nearly quintupled with the addition of the age parameter.

To determine the best model out of the two displayed in table 3, the model parameters can be evaluated. According to the coefficient of determination and the residual standard deviation, model m_1^d fits the data the best. Moreover, comparing the two models based on the difference between their

Table 3: Diagnostics of the two models fitting best to the data, where R^2_m is the marginal R^2, R^2_c is the conditional R^2, and $\hat{\sigma}$ is the residual standard deviation, or model sigma, measured in seconds.

Model	Fixed effect / Predictor	p-value[†]	Δ AIC[‡]	R^2_m	R^2_c	$\hat{\sigma}$ (s)	Error rate
m^d_1	maxLeftStanceTime	0.0538 .	-2.16	0.2442	0.9395	0.0765s	0.87%
	kneeRangeOfMotionDiff	0.0014 **					
	hipMidRangeDisbalance	0.0243 *					
	age	0.0020 *					
m^d_2	maxLeftStanceTime	0.0350 *	-2.43	0.2426	0.9362	0.0783s	0.89%
	kneeRangeOfMotionDiff	0.0021 **					
	hipHighRangeDisbalance	0.0467 *					
	age	0.00175 *					

[†] Significance codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < ' ' < 1
[‡] $\Delta AIC = AIC_i - AIC_{prev}$, where AIC_{prev} is the AIC obtained by the previously preferred model.

AIC values results in a slightly lower AIC for model m^d_1, strengthening the notion that it is the better model. The model diagnostics in table 3 further confirm this, as both the marginal and conditional R^2 are higher, and the model sigma and the error rate are lower.

3.5.3 Model Criticism

To ensure that the model can be correctly interpreted, the model assumptions must be checked. One of the assumptions mentioned in section 3.5.1 that require checking is about influential data points, which can be found by checking how much the model coefficients change when a data point is removed from the dataset. The DFBETAS are displayed in the left panels of figure 19, where the dashed lines represent the threshold of $2/\sqrt{n} = 2/\sqrt{80} \approx 0.2236$. Any observations with DFBETAS greater than the threshold are points of interest that are investigated using the panels on the right side of the figure. In the scatter plots of the fixed effects and the outcome variable, the participant IDs corresponding to the influential observations are marked. It should be noted that the range of measurements for all four fixed effects is within the limits of what is physically possible. For example, a maximum stance time of 4 seconds would indicate a measurement error as at no point in time (during the test) were the participants standing completely still. The highest value measured for this fixed effect is slightly longer than a quarter of a second, which, for a fast-paced change-of-direction test, is deemed reasonable. This is confirmed to be a common value for the predictor by assessing the face validity. The number of frames the foot visibly touches the ground is counted during a step. In most cases, this number counted up towards 10 frames, corresponding to a stance time of approximately 0.167 seconds (as the videos were recorded at 60 fps). No data points are removed due to the influential data points based on *maxLeftStanceTime*.

Next, the difference in knee range of motion is investigated. The marked observations in figure 19d are all inside or close to a relatively dense area of the scatter plot, meaning the influential data points do not correspond to any outliers. However, some of the other observations are far removed from the dense area but are not influential. To interpret this correctly, it's important to note that an outlier is not the same thing as an influential data point. This can be seen in figure 19d, where the highest value is an outlier but is not influential, and in figure 19f, where the highest value *is* an influential point. As the outliers for the difference in knee range of motion are not influential points, they are kept in the dataset. However, as the outlier marked with participant ID 45 in figure 19f is an influential data point, more information is required to determine the appropriate course of action.

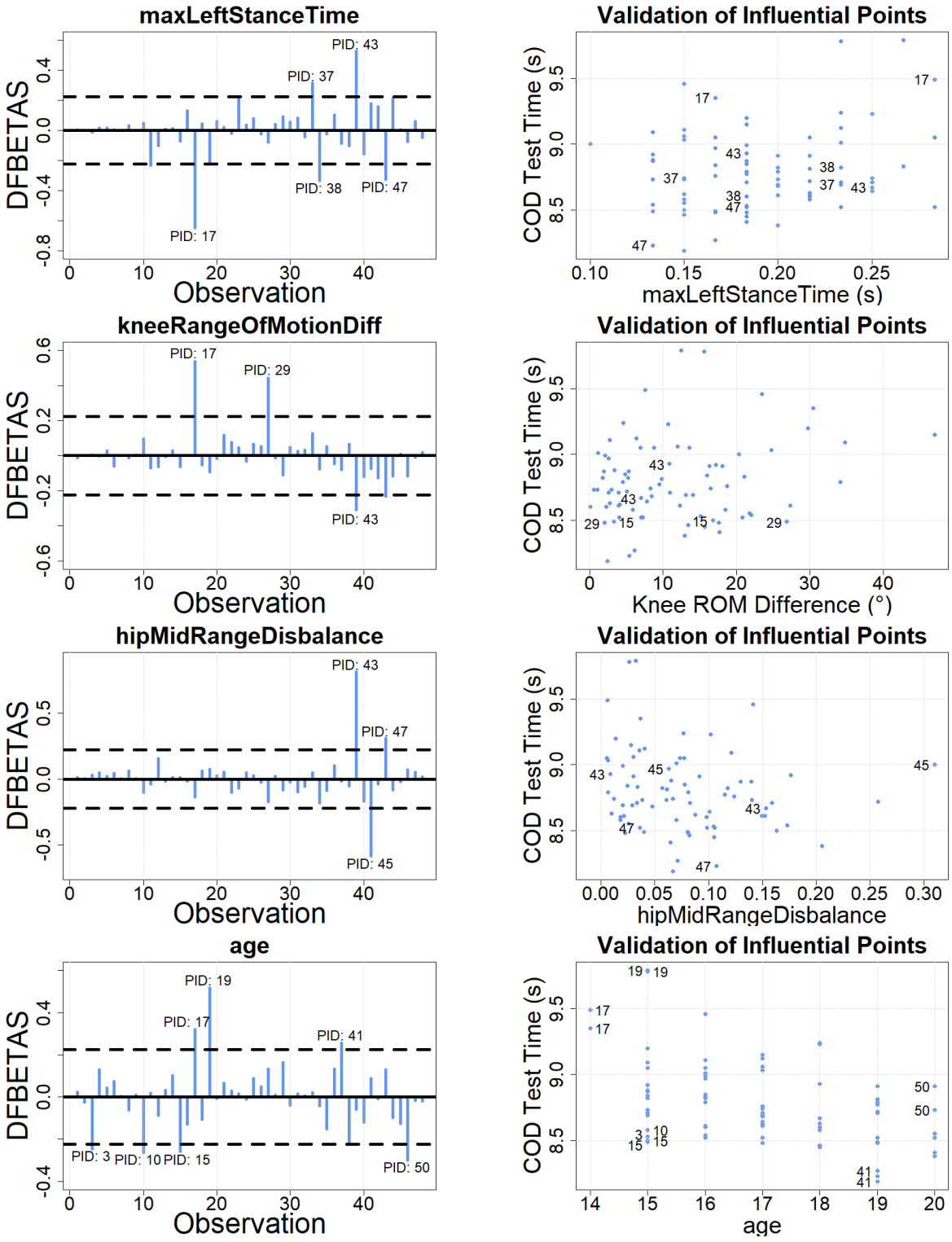


Figure 19: Left panels (a,c,e,g): DFBETAS values for each observation (where two tests of the same participant count as 1 observation) of model m_1^d. Right panels (b,d,f,g): Fixed effects plotted against the outcome variable CODTestTime

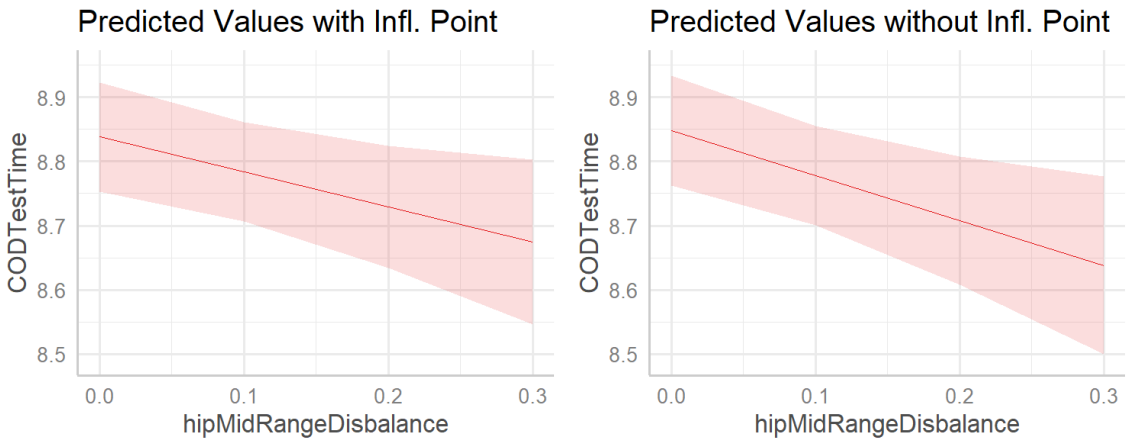


Figure 20: Model with (a) and without (b) the influential data point included regarding participant 45.

To understand the impact of the influential data point, the model can be compared to a model without the influential data point. As can be seen in figure 20, the influential data point changes the slope of the hipMidRangeDisbalance fixed effect only slightly (i.e. from approximately -0.546 to -0.698, as can be seen in table 4). By further examining the model coefficients, it is observed that removing the influential data point does not result in a sign change (i.e. positive estimates turning negative, and vice versa) for any of the fixed effects. Therefore, the interpretation of the model is not altered by removing the influential data point. Because of this, together with the observation that all data points have logical and realistic values, the influential data point is retained in the dataset.

Table 4: Difference in significance of model coefficients after removal of influential data point related to participant 45.

	Influental point included		**Influential point excluded**	
	Estimate	**p-value**[†]	**Estimate**	**p-value**[†]
(Intercept)	9.934	< 2e-16 ***	9.921	< 2e-16 ***
maxLeftStanceTime	0.651	0.054 .	0.700	0.040 *
kneeRangeOfMotionDiff	0.006	0.001 **	0.006	0.002 **
hipMidRangeDisbalance	-0.546	0.024 *	-0.698	0.011 *
age	-0.076	0.002 **	-0.075	0.002 **

[†] Significance codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < ' ' < 1

Linearity, Homoscedasticity & Collinearity
The other three assumptions can be confirmed by checking the residual plot and the VIF values. By observing the residual plot in figure 21, it can be seen that the data shows no obvious curved patterns that would signal a violation of the linearity assumption. Moreover, no clear increasing, decreasing, or modulating variance can be discerned from the residuals that would indicate heteroscedasticity. Therefore, the variability of the model data can be said to be linear and homoscedastic.

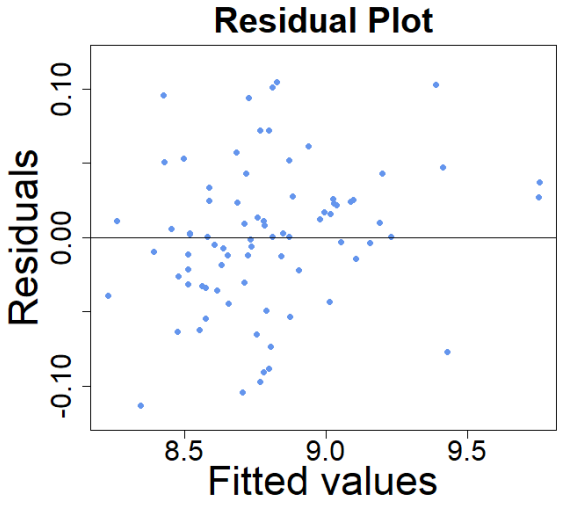


Figure 21: Residual plot of model m_1^d for checking linearity and homoscedasticity.

The remaining assumption can be checked with the VIF values displayed in table 5. The values for all four coefficients are close to 1, indicating that there is close to no collinearity present between the fixed effects. As James et al. (2013) mentions, there is always a small amount of collinearity among predictors, and the amount of collinearity present in the model is deemed not problematic.

Table 5: Variance Inflation Factors per fixed effect for model m_1^d.

Coefficient	**VIF**
maxLeftStanceTime	1.062
kneeRangeOfMotionDiff	1.153
hipMidRangeDisbalance	1.149
age	1.006

Chapter 4

Results & Discussion

In this chapter, the results relating to both research questions are described and simultaneously discussed. The applicability of OpenPose is discussed first (answering **RQ2**), which is followed by an evaluation of how biomechanical features were measured and what the results imply for practice (answering **RQ1**). The chapter concludes with the limitations of the presented work.

4.1 Applicability of OpenPose

Throughout the development of the data pipeline, in which a set of video recordings is transformed into tangible quantifications of the physical aspects of agility, observations and realisations have led to an understanding of the applicability of OpenPose in a fast-paced human movement scenario. However, proper validation of the 3D reconstructed movement data and the parameters derived from it has yet to be performed. This did not fit within the scope of this thesis and is something that can be done in future research building upon the work presented here. How the methods were justified and evaluated throughout the project is discussed in the following sections.

4.1.1 Camera Set-Up and Calibration

The pipeline starts with the camera set-up and calibration, which was tested before the actual data collection was planned. It was found that cameras should be placed perfectly horizontally, as, otherwise, the triangulated height of keypoints would incorrectly increase as they move away from the camera. Additionally, the ideal distance between cameras was determined by reviewing literature describing similar studies and by trying it out during a hands-on pilot test. The resulting between-camera distance seems to be a trade-off between capturing as much 3D information as possible while also keeping the calibration pattern in view of both cameras. It might be interesting to research other methods of stereo calibration that might not require a calibration pattern. This would allow for a wider angle between the two cameras and the test area, resulting in a smaller capture volume where the triangulated point can be placed, as mentioned by Zago et al. (2020). Essentially, the only thing that is calibrated in a stereo calibration of two cameras with known intrinsics is their relative positions in space. Detecting a recognizable object from both perspectives might be sufficient in placing the object in a shared three-dimensional space. However, this would most likely require significantly more manual development as this is not a method currently supported by MATLAB or OpenCV.

Another finding related to the camera calibration method was noticed when, at first, the intrinsic parameters and the relative positions of the two cameras were attempted to be calibrated simultaneously. This resulted in an unacceptably large reprojection error, which was corrected by splitting up the calibration. The intrinsic parameters of both cameras, retrieved from individual camera calibrations, could be used as input for the stereo calibration, resulting in a much lower and acceptable reprojection error. However, the error could most likely have been even lower if the calibration pattern were flatter. Additionally, the reflections of light on the calibration board might also have been disadvantageous for pattern detection. Corners in between black and white squares should have been clearly visible but appeared as one smudge of light instead. This would be less of a problem if the test was performed indoors, where the light is more static and the reflections are predictable. However, as change-of-direction tests performed by football players are rarely indoors, a matte calibration board would probably be a better option.

4.1.2 OpenPose Configuration

Next to flags required to save the detected keypoints, the OpenPose settings used included flags to improve its accuracy, like the *tracking* and *number_people_max* flags (see section 3.2.5 on page 31). However, there are other flags not investigated that might further improve OpenPose's performance. One example of this is the *net_resolution* flag, which, according to the OpenPose documentation (Hidalgo et al., n.d.[a]), potentially increases accuracy. The network resolution is a parameter that specifies the resolution at which the input image is processed by the neural network. The maximum speed-accuracy balance is said to be obtainable by using values closest to the aspect ratio of the to-be-processed images. However, as the default is recommended to let OpenPose find the ideal resolution (Hidalgo et al., n.d.[b]), the configuration used during this project might already have resulted in the most accurate pose detection. Concerning the *model_pose* flag, the BODY_25B model was chosen for its expected better performance and accuracy compared to its predecessor, the BODY_25 model. However, other models have not been actively investigated.

4.1.3 Data Pre-Processing

After the OpenPose data is triangulated to acquire 3D coordinates of every keypoint, the data seemed somewhat inconsistent. The length of limbs, for example, differs in every frame, although slightly in most cases (i.e. a few millimetres). This is an effect likely caused by inconsistent keypoint detection. For example, the distance between the hip and knee can differ when the knee is detected at the joint centre in one frame but in a more lateral or distal location from the joint centre in the next frame. Another cause might lie with incorrect triangulation. As the triangulation method makes use of the camera parameters acquired earlier in the process, part of the inconsistencies might already be solved by limiting the reprojection error in each calibration. However, if the keypoint detection remains inconsistent, then the triangulation will never be fully accurate. A common saying is that machine learning models are only as good as the data they are trained on. It is possible that OpenPose was not trained on images of COD or agility tests, causing the recordings of the arrowhead test to be the first time the algorithm has seen data like this. This might cause the model to have difficulty detecting poses in these recordings.

Addressing potential underlying causes of the scaling issue would be a difficult and time-consuming challenge. Therefore, the problem is mostly circumnavigated by adding an additional scaling step to the pipeline. Here, the thigh length is used to calculate a scaling factor that is used to resize all keypoints. The actual thigh length differs for each participant, but as an approximation, the length reported by de Leva (1996) is used. When looking at figure 12 on page 35, it can be observed that some running patterns seem larger than others. Although it is natural that some participants take relatively wide turns, the difference might be caused by scaling using a factor that is not completely tailored towards the unique bodily proportions of each participant. During scaling, the locations of all keypoints change positions with respect to the world origin. Therefore, the cameras might not be scaled properly for the tests of participants where the part of the running pattern between points B and C (see figure 2c on page 17) is closer to the cameras.

In turn, this most likely affects the extracted features that draw upon keypoint coordinates, such as the step length, take-off distance, velocity and acceleration. It does not affect the calculated joint angles, as the angles between keypoints remain the same when scaling all keypoints with the same factor. The scaling can potentially be corrected when the thigh lengths (or any other easily measurable length between two joints with corresponding OpenPose keypoints) are measured as part of the test. If this were to be done, then it is crucial to measure the distance between joints consistently so that the measured value can be realistically used to differentiate participants' bodily proportions. However, this method assumes that when two participants share the same thigh length, they also share the same height. In reality, this is not the case, as some people have relatively long legs compared to others. Whether or not the thigh length of each participant is measured is a trade-off between accuracy and pragmatism. The latter was preferred slightly over accuracy, as the accuracy obtained without the thigh measurements was considered sufficient for the purpose at hand.

As a final step, before features are extracted from the data, frames in which the participant is too far away from the camera or is occluded by an element of the test set-up are removed. This was a necessary step to acquire meaningful parameters. If players only had to make a turn with their feet going around the corner, then low pylons would suffice. However, as it is a requirement of the test for

players to run around each corner with their whole body, the high poles are required for the arrowhead test. It was known beforehand that these poles were part of the test set-up, but as OpenPose was expected to handle occlusion well, this did not seem a problem. However, it turned out to be a larger problem than anticipated. Although no literature was found that stated that other markerless motion capture methods were more capable of handling occlusion than OpenPose, it might be worth it to investigate this further. Additionally, pose detection might have been less difficult if a different COD test had been recorded. Using the arrowhead test was a deliberate choice, as it was found to be reliable in measuring COD in football players. However, with the current knowledge, the t-test, 5-0-5 test or the pro-agility test might have resulted in fewer issues. These tests merely require flat training cones, and additionally, the cameras can be placed much closer. The use of a different COD test does not stand in the way of extracting any of the features proposed in this thesis, as they all require the same type of movements from the participants.

4.1.4 Feature Extraction

During this project, five main features were extracted from the OpenPose data: joint angles, centre of mass, ground contact time, velocity, and acceleration. All parameters that are derived from the features are calculated based on the full observable part of the data (i.e. the orange part of figure 12 on page 35). This means that the parameters do not distinguish between clearly different segments of the test route, such as straight parts in between turns and the turns themselves. It might be worth investigating whether a segmented analysis of the test produces a different outcome.

However, the features inherently cover some parts of the test more than other parts. For example, the maximum step length, velocity and positive acceleration are related mostly to the straight sections, although they are calculated for the whole test. Likewise, the maximum negative acceleration is only relevant for the sections right before a turn, and the double support time is almost exclusively related to the turn sections. Therefore, some of the parameters will likely not change much when the analysis is split into segments. However, it might be good practice to remove certain segments from the calculations of parameters when it is known those segments do not contribute to the meaning of those parameters. As the performed analysis was not segmented, the following subsections will discuss the results of the feature extraction based on the full range of frames.

Joint Angles
The joint angles are calculated for the hip, knee and ankle joints, both left and right. It is difficult to validate the angles without comparing them to some "gold standard" (which can be done in future work using the data measured with the MVN Awinda system), the joint angles were checked for accuracy using a face-validity approach. As can be seen in figure 22, the 3D reconstructed pose, the angle in graph form, and the corresponding video frames were assessed for multiple tests and participants. No incorrect angles have been observed throughout this process.

Although the range of motion and the disbalance parameter both seem to have a correlation with the test completion time (as can be seen in figure 23 on page 56), one important caveat must be noted. In the current work, it is assumed that each angle that is measured is the product of flexion and extension. For example, a healthy knee will mostly bend in the sagittal plane about the frontal axis. However, for some people more than others, the knee can also bend in the frontal plane about the sagittal axis. This is referred to as knee abduction or adduction, and, as already mentioned in section 2.2.3 on page 19, this can lead to an increased risk of knee injuries. Therefore, it is important to be able to distinguish in which plane a knee is rotated. Sensor-based motion capture often includes the principal axis of every segment, which makes it possible to calculate angles relative to neighbouring segments. However, using only OpenPose data, which does not come with information about the orientation of body segments, this is difficult to achieve.

Additionally, the range of motion calculation does not allow for angles greater than 180°. This means that when joints are hyperextended (which does not occur often in a COD test), a knee angle of, for example, 185° would be measured as a 5° angle. Such a small angle is filtered out by the *rmoutliers* function, which removes data more than three scaled median absolute deviations from the median, and therefore regards potentially interesting information as incorrect.

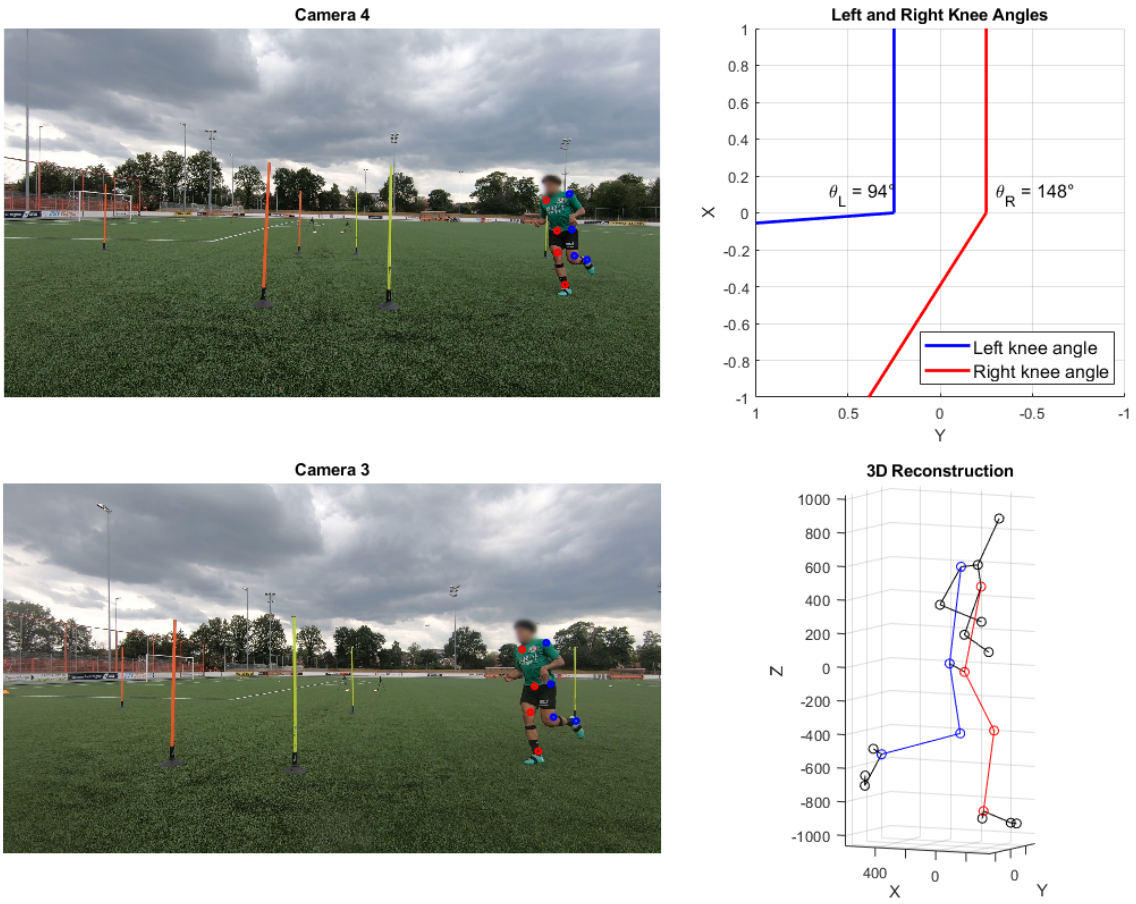


Figure 22: Screenshot of the face validity assessment of the joint angles.

Centre Of Mass

The centre of mass has been calculated by using the segment masses and relative segment COM positions proposed by de Leva (1996). To achieve the desired outCoMes, the segment endpoints of de Leva (1996) are matched with keypoint locations. Some endpoints, like the upper arms, forearms, thighs and shanks, correspond directly to OpenPose keypoints. However, some do not have a direct match, meaning the endpoints needed to be as closely approximated to a keypoint as possible. For these segments, i.e. the head, trunk and feet, the segment masses and relative COM position along the segment's principal axis might not be fully accurate. This, in turn, could lead to a slight deviation of the calculated COM position compared to the actual COM position. To accommodate for any inconsistencies present in the data due to this, or due to incorrectly detected keypoints, the data is filtered. Any outliers are replaced with linearly interpolated values, and a Butterworth filter is applied to remove any remaining noise in the data. As could be observed in figure 13 on page 38, both steps were necessary to achieve proper results.

For essentially the same reason as with the joint angles, the COM was validated through face validity. This was done by plotting the centre of mass position together with the 3D reconstructed pose for each frame of a video (an example is shown in figure 14 on page 39). It is known that the COM of the human body lies slightly above the hips, and depending on a person's pose, the COM can be positioned outside of the human body. With this in mind, the calculated locations were assessed and found to be in logical positions relative to the body.

Ground Contact Time

The ground contact time was calculated by determining, for each step, the moment the feet first touch the ground and the moment they leave the ground. These heel strike and toe-off events are based on peaks in the vertical velocity of the heel and toe keypoints. Like in previous evaluations, the face validity was assessed by comparing the original videos with the timestamps of the events. It is difficult to pinpoint these events precisely when the cleats (in Dutch: "noppen") of the shoes

disappear into the grass. However, to ensure the most consistency in the evaluation, the heel strike is determined to be the moment the heel does not "sink" further into the ground. The results seem to be mostly correct event times with only slight deviations. The deviations found were only about three frames apart from the observed events, corresponding to a mere 50 milliseconds off. However, it was also observed that the further the participant is away from the camera, the higher the deviation. This once again shows the importance of the percentage of the image area the participant occupies for achieving accurate pose detection.

Another point of discussion is related to the applicability of heel strikes and toe-offs in scenarios other than gait analysis. When a person is walking, the heel always strikes the ground first, and the toes are always the last part of the foot to leave the ground. However, when running in a COD test (but also in sprint tests), the heels do not even touch the ground at (near-)full speed. Therefore, for the purpose of this study, the heel strike might not be the best fit for the start of the ground contact time. However, the vertical velocity of the heel is measured, which will also have a negative peak very near the first ground contact. Therefore, it is not expected to result in a drastically different ground contact time when this is taken into account.

Velocity And Acceleration
The velocity and acceleration could be calculated using the position of the centre of mass over time. To assess whether the calculations did not result in any values outside the boundaries of what is physically possible, the data was plotted and compared to reference values. As can be seen in figure 17 on page 42, most of the velocities and accelerations lie within these boundaries, and the exceptions are only slightly over the reference value. These might be due to a difference in participants compared to Postma et al. (2022). They recruited healthy, injury-free individuals aged 19-22, whereas, for the present study, all participants were experienced youth football players of a professional football academy aged 14-20. It is possible that the participants of the present study, who practise several times a week, were fitter and thus faster than the participants of Postma et al. (2022). One could wonder whether a higher acceleration can be reached in a sprint test compared to a change of direction test, as the linear distances covered are longer. However, the model shows that the maximum acceleration is reached within the first second of accelerating. This means that the same acceleration can be reached during both tests. The opposite is true for the velocity, for which it takes a little longer to reach its maximum. According to Postma et al. (2022)'s model, speeds of 9 ms^{-1} can be reached during a sprint. However, their sprint distances were at least 7.5 m and the maximum distance covered during the observable part of the COD test was only 5 m. Therefore, the participants of the arrowhead test might not have been able to reach the top speeds the participants of Postma et al. (2022) have reached.

4.1.5 Concluding Remarks

One of the research questions (**RQ2**) set to be answered by this project was: *To what extent is markerless motion capture suitable for measuring biomechanical features of agility?* In this section, the method developed to answer this question has been discussed. It became clear that it is possible to derive certain biomechanical features from stereo videos that are meaningful for movement analysis. However, there are also a few impediments throughout the process that influence the suitability of the methods used. The velocity and acceleration, as well as the centre-of-mass positions over time they are based on, lend themselves rather well to be measured using MMC (with the necessary data processing). However, the joint angles can only be calculated regardless of the anatomical plane in which the rotation takes place. For the ground contact time, face validity assessment determined the times to be sufficiently accurate. However, as is true for all extracted features, without proper validation through comparison with a gold standard, no definitive conclusions can be given about the method's accuracy.

4.2 Measuring Biomechanical Features of Agility

The method described in the previous section has been the result of an investigation of literature into how agility and change-of-direction speed can be measured. The research resulted in a set of overarching biomechanical features that can essentially be measured from any stereo video recording of a person performing change-of-direction movements. All parameters derived from the biomechanical features were analysed using a Generalized Linear Mixed-Effect Modelling approach. Parameters were added to models in an iterative manner to find the ones that could be significantly associated with agility. This resulted in a model consisting of four fixed effects and one random effect. However, because the data has not been compared to a gold standard yet, the analysis and implications of the data, as described in this section, can only serve as a demonstration of how the data can be interpreted after the data has been validated. Because of this, no definitive conclusions can be given about the features' significance to agility.

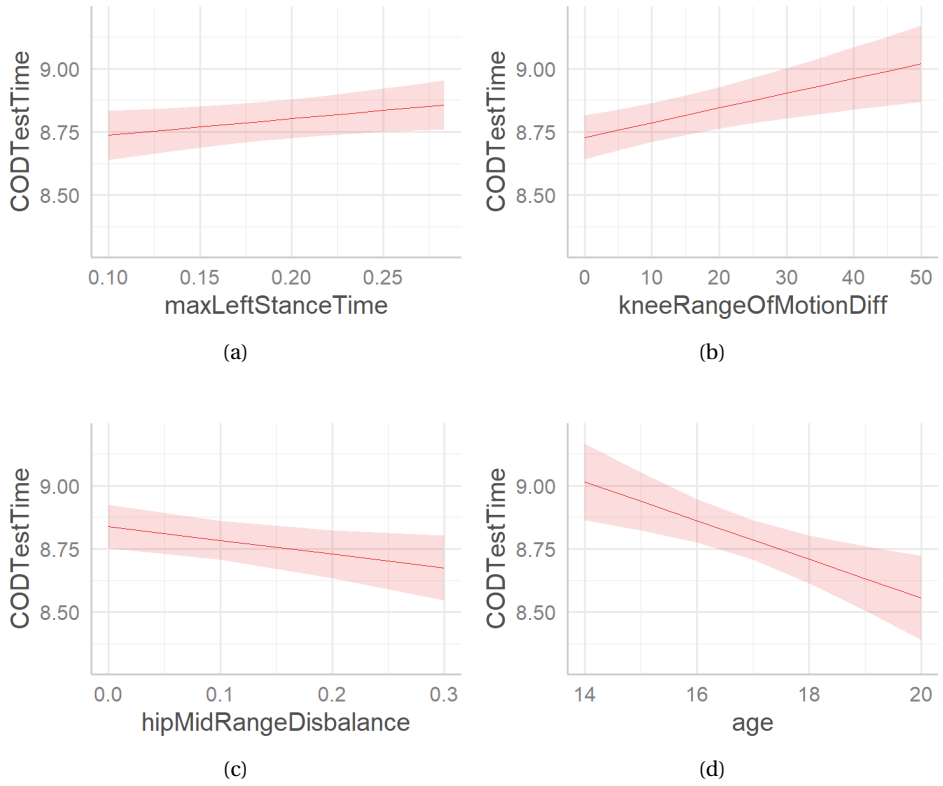


Figure 23: Predicted linear relationships between fixed effects and the outcome variable according to model m_1^d.

4.2.1 Statistical Analysis and Model Evaluation

To determine the best model out of the two displayed in table 3 on page 47, the model parameters can be evaluated. According to the coefficient of determination and the residual standard deviation, model m_1^d fits the data the best. Moreover, comparing the two models based on the difference between their AIC values resulted in a slightly lower AIC for model m_1^d, strengthening the notion that it is the better model. The model diagnostics in table 3 further confirm this, as both the marginal and conditional R^2 are higher, and the model sigma and the error rate are lower. The linear relationships between the fixed effects and the outcome variable of the model are shown in figure 23. While this all points in the direction of m_1^d, the maxLeftStanceTime fixed effect only has a p-value of 0.0538, which is above the significance threshold of $p < 0.05$. This might indicate that the correlation between the fixed effect and the test completion time is not strong enough.

Table 6: Significance of likelihood ratio tests when removing each of the fixed effects.

Full model *CODTestTime ~ maxLeftStanceTime + kneeRangeOfMotionDiff + hipHighRangeDisbalance + age + (1|PID)*	
Null model	**p-value**[†]
CODTestTime ~ kneeRangeOfMotionDiff + hipHighRangeDisbalance + age + (1 | PID)	0.04208 *
CODTestTime ~ maxLeftStanceTime + hipHighRangeDisbalance + age + (1 | PID)	0.00078 ***
CODTestTime ~ maxLeftStanceTime + kneeRangeOfMotionDiff + age + (1 | PID)	0.01742 *
CODTestTime ~ maxLeftStanceTime + kneeRangeOfMotionDiff + hipHighRangeDisbalance + (1 | PID)	0.00147 **

[†] Significance codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < ' ' < 1

To check the extent to which it contributes to the overall performance of the model, a likelihood ratio test (LRT) is performed. As can be seen in table 6, the difference between the null model not containing maxLeftStanceTime and the full model is significant ($p < 0.05$). Using the same tests, the other fixed effects can also be evaluated. This is relevant, as the differences in total explained variance between models m_1^d and the intercept-only model do not differ greatly. This begs the question of whether the addition of the fixed effects significantly improves the model fit to the data compared to a model with just a random effect. Therefore, LRTs are additionally performed for each of the other fixed effects of model m_1^d. As can be seen in table 6, the differences between the full model and each null model are all significant ($p < 0.05$). The kneeRangeOfMotionDiff is especially significant with $p < 0.001$.

Table 7: Model summaries of included fixed effects.

Original Fixed Effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	9.934	0.403	49.954	24.65	< 2e-16 ***
maxLeftStanceTime	0.651	0.326	34.028	1.998	0.054 .
kneeRangeOfMotionDiff	0.006	0.002	38.654	3.439	0.001 **
hipMidRangeDisbalance	-0.546	0.232	35.847	-2.351	0.024 *
age	-0.076	0.023	46.164	-3.274	0.002 **

Standardized Fixed Effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.796	0.039	45.862	228.363	< 2e-16 ***
maxLeftStanceTime.z	0.027	0.014	34.028	1.998	0.054 .
kneeRangeOfMotionDiff.z	0.055	0.016	38.654	3.439	0.001 **
hipMidRangeDisbalance.z	-0.032	0.014	35.847	-2.351	0.024 *
age.z	-0.128	0.039	46.164	-3.274	0.002 **

[†] Significance codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < ' ' < 1

Model m_1^d is summarized in table 7. Additionally, the model with the fixed effects centred and scaled is displayed. Standardizing parameters makes sure that the coefficients can be meaningfully compared to one another by transforming them to the same scale (James et al., 2013). The standardized slopes, displayed in the 'Estimate' column in the lower half of the table, are an indication of the effect sizes of the fixed effects and therefore show how strongly the parameters are associated with the outcome variable. In the lower part of the table, it can be observed that the association with the age of the participants is the strongest based on the steepness of its slope. Of the parameters derived from OpenPose, the difference in knee range of motion is most strongly associated with the test completion time. This corresponds with the significance of the parameter in the LRT mentioned earlier.

One potential issue that needs to be addressed regards the dataset. Some participants only appear once in the dataset due to technical errors that occurred during data collection. However, in GLMMs, the random effect is introduced to let the model know that there are multiple observations for each value of the random effect (in this case, the participant ID). This might give rise to an identifiability issue for participants with only a single observation. In equation 14 on page 43, it can be argued that when there is only one observation X_i for random effect Z_j, then the model cannot properly distinguish between Z_j and the residual variance ϵ_j as the sum of the two remains constant. This means that single-observation random effects do not contribute to estimating the overall participant random effect variance of the model. However, they can still be used to estimate the mean structure of the outcome variable. Therefore, the single-observation participants are not excluded from the dataset.

4.2.2 Model Implications

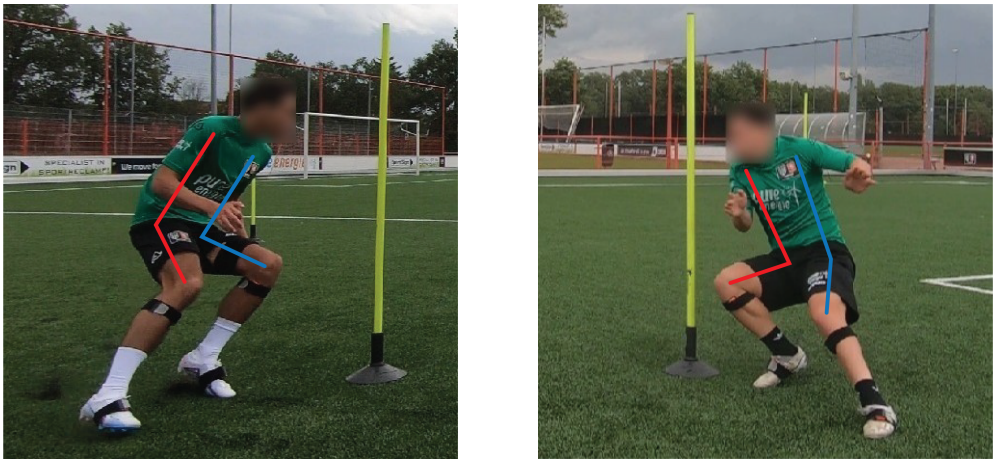
To understand what the statistical analysis means, the implications of the model are discussed. It was found that the combination of the maximum left stance time (related to GCT), the difference in left and right range of motion of the knee (related to asymmetry), the disbalance in the hips in the range of 60°–120°(also related to asymmetry), and the age of the participant resulted in the model with the highest explained variance. The conditional R^2 is 0.9395, meaning that the variance explained by both fixed effects and random effects is 93.95%. Interestingly, the value is not much higher than the conditional R^2 achieved using only a random intercept per participant in the model. This indicates that a substantial portion of the variability in the test completion time is attributed to individual participants. To some extent, this was to be expected, as each participant's test times were very close to each other. That would imply that the range of outcome values per participant is limited to only a fraction of the range of outcomes of the full dataset. However, this should also be seen as the strength of a mixed model, as the method excels in personalized predictions. Using the model and a known participant ID, a trainer can make very refined predictions about how a change in an independent variable will affect that particular person. Examples are given in the following discussions per parameter of the best-fitting model found.

While the small difference in R_c^2 between the best-found model and the intercept-only model might indicate that the fixed effects do not contribute much to the overall performance of the model, they do contribute to understanding what predictor parameters are significantly associated with CODS (assuming the data is correct). The reason is that the marginal R^2 has a value of 0.2442, meaning that 24.42% of the total variance in completion time is explained by the fixed effects. A large portion of this is due to the age of the participant, as the marginal R^2 before the addition of the parameter was only approximately 5%. Although it is no biomechanical parameter, it is a useful insight as it can be used by trainers of specific age-grouped teams. Without the age taken into account, a player might not appear to be very agile, but perhaps for their age, they are. Coaches can have their players train specifically to improve their agility if it is known that they underperform compared to other players their age. The reason why the age is significant is unknown. However, possible explanations might be associated with the branches of CODS in figure 1 on page 15. For example, older players have had more training hours to improve their technique. Another reason might be that their muscles are more developed, allowing them to have greater concentric strength and power. Further research is required to learn about potential contributing factors of agility related to age and perhaps other anthropometric variables.

The age can also be used to predict how a player is expected to perform over time. Figure 23d on page 56 shows that when a participant ages two years, he is expected to execute the COD test ap-

proximately 0.125 seconds faster than before. Coaches can use this to gauge whether a player is performing to their full potential. Furthermore, the age parameter poses an interesting discussion. Say a player sets the fastest test time in his age group at a certain moment in time, but five years later, he is the slowest in their age group while still completing the test quicker than he did five years earlier. He might be more agile than he was, but compared to his peers, he is rather slow. In such a case, it might be good to check his and his peers' data to see whether this person's CODS has actually stagnated or that his peers have improved relatively greatly. Additionally, the differences in test times might have become smaller compared to five years ago. It is essential that data is collected regularly to allow coaches to see how much players have progressed over the years and to make comparisons over time and between players possible.

As for the other fixed effects, it seems that interlimb asymmetry, especially for the knees and the hips, is associated with COD ability. However, the relationship between COD ability and the hip disbalance parameter is contrary to what was expected. As was found by Bishop et al. (2021) and Philipp et al. (2021), a disbalance between left and right limbs should indicate a lowered COD performance, but the fixed effect shows a shorter completion time for greater imbalances. This might have to do with the way the parameter is calculated. When speaking of the hip angle, the angle between the trunk and the thigh is meant, and the mid-range angles are those between 60° and 120°. A mid-range disbalance of the hip means that there is a difference between the total duration in which the left and right hips were in that range of angles. Therefore, a "disbalance", in this case, might be an indication of how a participant turns around a corner, i.e. a tactical factor rather than biomechanical. This is illustrated in figure 24, where figure 24a shows an example of a participant with a low disbalance, and figure 24b one with a high disbalance. What can be seen is that, when turning around a corner, the placement of the outer foot is more distal for participants with a high hip mid-range disbalance. Furthermore, the knee of the outer leg is further extended for these participants, causing a larger difference between the left and right hip angles. A greater difference in the angle might correspond with a low centre of mass, which might benefit an efficient change of direction. However, this has yet to be confirmed.



(a) Participant with low disbalance (b) Participant with high disbalance

Figure 24: Examples of a participant with a low disbalance value (a) and a high disbalance value (b). The difference between the left and right hip angles is smaller for the participant in (a).

The increased performance might also be due to greater braking forces associated with more distal foot placement, but as this has not been researched during this thesis, this presumption can only be confirmed by further research. Nonetheless, trainers are recommended to steer their underperforming players into improving their posture when rapidly changing direction by keeping the centre of mass low to the ground while extending the outer leg (compared to the contralateral leg) in the steps close to the turn.

The other asymmetry-related predictor included in the model is the difference between the range of motion of the left and right knees. This parameter is, apart from age, the most significant parameter according to its effect size in table 7 on page 57. When the difference in range of motion between the knees is smaller, the test completion time is shorter, and the COD performance is, therefore,

greater (as can be seen in figure 23b on page 56). As opposed to the hip mid-range disbalance, this does concur with the findings of Bishop et al. (2021) and Philipp et al. (2021). The result implies that when test participants use the full range of motion of both knees (or at least to a greater extent), they are able to change directions quicker as compared to participants who only use part of their range of motion. More specifically, players are expected to perform the arrowhead test 0.05 seconds faster when their knee range of motion difference is decreased by 10°. This might not seem much, but it might, at the same time, be something that is easily improved. If, for example, a player is not using his full range of motion of a certain leg, causing a difference of 30°, then that player "only" has to improve the range of motion of his lesser leg to the level of his other leg to decrease his test time with 0.15 seconds. However, it is most likely that there are other factors that contribute to this difference that are more difficult to tackle.

Putting this next to the implications of the disbalance parameter, the ROM poses an interesting discussion. On the one hand, participants are thought to perform better when their difference in hip angles is large during a turn. Yet, on the other hand, the difference in the range of motion of the knees should remain small. Investigating the role of these angles (i.e. of the hip and knee) relative to each other might result in valuable insights. A possible result of such a study might be that the efficiency of changing direction is related to how far up the knees are raised or, alternatively, how far down the trunk is lowered. This would again connect knee and hip angles to the centre of mass, which would feed into the assumption that CODS is all about one's ability to control their centre of mass. Additionally, the trunk is tilted forward (i.e. smaller hip angle) mostly when accelerating and straighter (i.e. larger hip angle) when at full speed or during deceleration. Therefore, this suggested study might also work well in combination with the segmented analysis proposed earlier on in section 4.1.4 on page 53. The mechanism behind efficiently controlling one's centre of mass likely varies between different sections of a COD test. As mentioned earlier, lowering it during deceleration might improve efficiency. Likewise, aligning the centre of mass and the trail leg at take-off with the direction one wants to go (i.e. having a significantly forward-tilted pose) might efficiently push the ground reaction force in the desired direction, allowing for increased acceleration efficiency.

Furthermore, the ground contact time association that was found by Welch et al. (2019), Dos'Santos et al. (2017), and Condello et al. (2016) was confirmed. The stance time parameter indicates that having a shorter maximum stance time of the left leg can be associated with decreased performance in the COD test. More specifically, a decrease of 0.15 seconds in the maximum left stance time is expected to decrease the test time by 0.1 seconds (see figure 23a on page 56). Thus, when players can manage to keep the time their feet touch the ground for each step to a minimum, they are able to change directions faster. Why only the parameter for the left leg came forward as significant might have to do with the preferred foot of the participants (besides data being potentially inaccurate). It could be the case that most participants preferred their right foot (as the majority of the world is right-handed), and there is some correlation between stance time, turn direction and "footedness". However, foot preference information could not be obtained in time, so this was not included in the model parameters. It might prove useful to investigate this further, as it could be the case that the faster test completion times were obtained by participants who were (relatively more) ambipedal (i.e. those who can use both feet equally well).

There is still a significant portion of the variance not accounted for by the variables included. The additional unexplained variance could come from other biomechanical parameters not explored here, such as the centre of mass height during an acceleration, ground reaction forces in key moments of the test, or the strength and power of individual participants. There is much to explore in this field, of which some recommendations for future work can be found in section 5.1 on page 64.

4.2.3 Concluding Remarks

The research question left to be answered (**RQ1**) is: *How can biomechanical features of agility be measured in a practical manner beyond the confines of a lab setting?* Answering this question started off with diving into literature with the goal of finding biomechanical associations with change of direction speed and ability. Three main factors were found: interlimb asymmetry, ground contact time (GCT), and take-off distance. Using OpenPose, three-dimensional coordinates of 20 anatomical landmarks (i.e. keypoints) were derived from a set of video recordings, which were used to cal-

culate the biomechanical features. Based on this, 65 variables that could theoretically play a role in assessing a person's COD ability were calculated. Additionally, seven variables related to speed or acceleration and another four variables related to demography and anthropometry were added to the dataset. Each of these variables was analyzed by using them as fixed effects in Generalized Linear Mixed-Effects Models. As stated earlier, without comparing the data to a gold standard, definitive conclusions cannot be drawn. However, the biomechanical parameters can all be measured in a practical manner using only a set of cameras. Although the parameters are currently calculated from recordings of a COD test, which is a lab setting, they can essentially also be calculated with video recordings of regular football practice.

Chapter 5

Conclusion

This thesis has served as a first step towards analysing agility using markerless motion capture. Most research using markerless motion capture is conducted for gait analysis, which is often a two-dimensional and one-directional assessment. However, an agility test contains multidirectional movement, requiring three-dimensional movement data. To address this, a data pipeline was developed in which biomechanical features relevant to agility are derived from stereo video recordings of participants in the dynamic, fast-paced environment of a change-of-direction test.

The research first set out to answer the research question: **How can biomechanical features of agility be measured in a practical manner beyond the confines of a lab setting?** Through literature research, correlations were found between biomechanical features and the physical component of agility (i.e. not related to cognition). Interlimb asymmetry, ground contact time, and the take-off distance were included in the pipeline. These were expected to be measurable using MMC, as opposed to, for example, ground reaction forces. Calculating parameters for the selected biomechanical features required joint angles, the centre of mass, and foot-ground contact start and end times for each step to be computed. Additionally, the velocity and acceleration of the participants were calculated. The parameters could be measured in a way that did not take extra time per participant. While the parameters were calculated from change-of-direction tests, they are not strictly tied to such tests and can hypothetically also be measured during regular football practice. Whether this is achievable yet, the answer to the second research question is needed.

The second research question was: **To what extent is markerless motion capture suitable for measuring biomechanical features of agility?** It can be concluded that by using OpenPose, it is possible to extract relevant information regarding agility. The technology offers a great deal of possibilities for the extraction of biomechanical data in a practical manner using off-the-shelf tools. OpenPose was relatively consistent for a considerable part. However, OpenPose has displayed a decreased pose detection accuracy when the person is more than approximately 10 meters away from the camera using a resolution of 2.7K. Additionally, when elements of the test set-up partly occluded the body, OpenPose could not always accurately detect the human pose. To acquire truly useful data, the pose detection must be more accurate and consistent. Additionally, the triangulation of the set of 2D OpenPose coordinates into 3D coordinates has been proven difficult to do accurately. When the reprojection errors within the camera calibration process are too large, the triangulation will not result in actual world distances regardless of the accuracy of the keypoint detection by OpenPose. Therefore, OpenPose is suitable for measuring biomechanical features of agility, but only once its pose detection accuracy is improved and the triangulation is done correctly.

To know the accuracy is sufficient with adequate certainty, the data must be compared to a gold standard. This has not been part of this thesis, essentially meaning that the validity of the data cannot be confirmed, although face validity assessments did not result in any apparent errors. For the same reason, no definitive conclusions can be drawn from the model predicting agility. However, age was not a feature calculated from OpenPose data but did turn up as a significant predictor. Therefore, it can be concluded that with age, people tend to perform better on COD tests. The reason for this might be related to the time older players have had to improve their technique and strength. However, this has yet to be researched.

5.1 Future Work

Some unexplored research directions can be considered to build upon, improve, and extend this work. The most important recommendation is to conduct a validation of the data using a gold-standard motion capture system. This evaluation can be conducted in work building upon this thesis because all the tests of which video recordings were made were also recorded using the MVN Awinda sensor-based motion capture system. The comparison is useful to evaluate the accuracy of the triangulated three-dimensional data that is used as the basis for calculating all biomechanical features in this thesis.

The accuracy of the developed method can presumably be improved by altering the extra scaling step that was introduced to the process after triangulation. Scaling the data based on a body segment measurement per participant, such as the thigh length, is likely to result in more accurate distances and body segment lengths.

Higher accuracy and consistency might also be attainable by slightly changing some aspects of the camera calibration and data processing steps. For example, using a calibration pattern printed on a thick and robust panel might decrease reprojection errors while calibrating the camera. Additionally, the accuracy of the pose detection might be improved by training the system on data similar to change-of-direction tests. This might make OpenPose more familiar with the video recordings that were used in this thesis. In turn, this might solve the occlusion and misdetection problem observed close to the test poles.

Furthermore, being too far removed from the camera resulted in decreased pose detection accuracy. The issue, as well as the occlusion problem, might be bypassed by using a different COD test than the arrowhead test. It is, therefore, recommended to use the t-test, 5-0-5 test or pro-agility test to allow the camera to be positioned closer to the test area. The tests also do not require the high poles that occlude the participant during the test.

Apart from the test set-up and data processing steps, some more research directions are recommended to explore. First and foremost, it is recommended to analyse the data in segments. For certain parameters, such as the maximum negative acceleration, only specific sections of the COD test are relevant. Two people might have a similar test time, but one might be faster in the straight sections, and the other might be able to turn around corners more efficiently. The latter might also be due to the participant's foot preference during turns in certain directions. Therefore, including information about whether a participant is left-footed, right-footed, or ambipedal might result in interesting insights.

Another aspect worth investigating relevant to a segmented analysis is the vertical position of the centre of mass. As stated in section 2.2.3, a lower centre of mass while (re-)accelerating, e.g. after a turn, was associated with cutting performance. In fact, it might be the case that being agile is all about how well one can control their centre-of-mass position when moving around. Therefore, it is also recommended that the role of centre-of-mass control is further investigated, perhaps related to relative foot placement and acceleration.

Furthermore, it might prove useful to extend this research by adding sensors to the set-up. However, practicality should remain a requirement. Sensor fusion should improve the accuracy, consistency, and/or reliability of the data while keeping any additional time the data collection costs per participant to a strict minimum. One could think of a minimal sensor set-up in which GPS data is used to calculate position and velocity more accurately. FC Twente, and likely also other professional football clubs, already use sports vests that include such trackers. These can be leveraged to gain additional insights on top of the MMC data. The data can also be used for research into whether models can be created that are more reliable than the one developed here as a demonstration. This might result in models that are better able to explain the variance present in the test completion time.

Lastly, it might prove useful to present the camera footage to trainers and sports scientists in combination with the extracted features. At the very beginning of this thesis, sports scientists of FC Twente and the FC Twente/Heracles Academie were already asked what they pay attention to during an agility test. While this did not result in any tangible insights, perhaps showing them the data with corresponding test images provides new insights. Additionally, comparisons between the superior-

and inferior-performing athletes might aid in understanding the difference in agility. An interactive dashboard can be created to present all the information. The specific wants and needs of the end users of such a display of data (e.g. trainers, coaches, and sports scientists) should be researched. However, it would be interesting to include interactivity surrounding the predictors, so the user can, for example, adjust the age of a player of interest to see how that player is expected to develop as he ages.

References

Adobe (n.d.). *Adobe Premiere Pro*. URL: https://www.adobe.com/products/premiere.html.

Akaike, H. (1992). "Information Theory and an Extension of the Maximum Likelihood Principle". In: *Breakthroughs in Statistics*. Springer Science and Business Media Deutschland GmbH, pp. 610–624. DOI: 10.1007/978-1-4612-0919-5{_}38.

Bai, Linge and Breen, David (Jan. 2008). "Calculating Center of Mass in an Unbounded 2D Environment". In: *Journal of Graphics Tools* 13.4, pp. 53–60. DOI: 10.1080/2151237X.2008.10129266.

Bates, Douglas, Mächler, Martin, Bolker, Ben, et al. (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1. DOI: 10.18637/jss.v067.i01.

Bazarevsky, Valentin, Grishchenko, Ivan, Raveendran, Karthik, et al. (June 2020). "BlazePose: On-device Real-time Body Pose tracking". In: *arXiv*.

Beckett, Claire, Eriksson, Lennart, Johansson, Erik, et al. (2017). *Multivariate Data Analysis (MVDA)*. 7th. Pearson Education Limited, pp. 201–225. DOI: 10.1002/9781118895238.ch8.

Belsley, David a, Kuh, Edwin, and Welsch, Roy E (June 1980). *Regression Diagnostics: Identifying influential data and sources of collinearity*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., p. 32. DOI: 10.1002/0471725153.

Bill, Kevin, Mai, Patrick, Willwacher, Steffen, et al. (Sept. 2022). "Athletes with high knee abduction moments show increased vertical center of mass excursions and knee valgus angles across sport-specific fake-and-cut tasks of different complexities". In: *Frontiers in Sports and Active Living* 4. DOI: 10.3389/fspor.2022.983889.

Bishop, Chris, Read, Paul, Brazier, Jon, et al. (Aug. 2021). "Effects of Interlimb Asymmetries on Acceleration and Change of Direction Speed: A Between-Sport Comparison of Professional Soccer and Cricket Athletes". In: *Journal of Strength and Conditioning Research* 35.8, pp. 2095–2101. DOI: 10.1519/JSC.0000000000003135.

Burnham, Kenneth P. and Anderson, David R. (2004). *Model Selection and Multimodel Inference*. Vol. 9780521887. New York, NY: Springer New York, pp. 79–98. DOI: 10.1007/b97636.

calib.io (n.d.). *Camera Calibration Pattern Generator*. URL: https://calib.io/pages/camera-calibration-pattern-generator.

Cao, Zhe, Hidalgo, Gines, Simon, Tomas, et al. (Jan. 2021). "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1, pp. 172–186. DOI: 10.1109/TPAMI.2019.2929257.

Cao, Zhe, Simon, Tomas, Wei, Shih-En, et al. (Nov. 2016). "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields". In: *arXiv*.

Cavanaugh, Joseph E and Neath, Andrew A (May 2019). "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements". In: *WIREs Computational Statistics* 11.3, e1460. DOI: 10.1002/wics.1460.

Chalil, Dudung Hasanudin, Febrianty, Mona Fiametta, and Sartono, Hadi (2017). "The Validity and Reliability of Arrowhead Agility Test in Football". In: *2nd International Conference on Sports Science, Health and Physical Education*. Vol. 1. SCITEPRESS - Science and Technology Publications, pp. 414–417. DOI: 10.5220/0007062104140417.

Condello, Giancarlo, Kernozek, Thomas W., Tessitore, Antonio, et al. (Jan. 2016). "Biomechanical Analysis of a Change-of-Direction Task in College Soccer Players". In: *International Journal of Sports Physiology and Performance* 11.1, pp. 96–101. DOI: 10.1123/ijspp.2014-0458.

Cudejko, Tomasz, Button, Kate, and Al-Amri, Mohammad (Aug. 2022). "Validity and reliability of accelerations and orientations measured using wearable sensors during functional activities". In: *Scientific Reports* 12.1, p. 14619. DOI: 10.1038/s41598-022-18845-x.

Dann, Eleanor, Quinn, Samuel, Russell, Mark, et al. (Nov. 2022). "Alternate Leg Bounding Acutely Improves Change of Direction Performance in Women's Team Sports Players Irrespective of Ground Type". In: *Journal of Strength and Conditioning Research* Publish Ah, pp. 27–31. DOI: 10.1519/JSC.0000000000004378.

Davis, Kenneth E., Ritter, Merrill A., Berend, Michael E., et al. (2007). "The importance of range of motion after total hip arthroplasty". In: *Clinical Orthopaedics and Related Research* 465.465, pp. 180–184. DOI: 10.1097/BLO.0b013e31815c5a64.

Daykin, Caomham (n.d.). *Maximum Acceleration and Deceleration – Metric Considerations and Uses - STATSports*. URL: https://pro.statsports.com/maximum-acceleration-and-deceleration-metric-considerations-and-uses/.

De Leva, Paolo (Sept. 1996). "Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters". In: *Journal of Biomechanics* 29.9, pp. 1223–1230. DOI: 10.1016/0021-9290(95)00178-6.

Dos'Santos, Thomas, Thomas, Christopher, Jones, Paul A., et al. (Mar. 2017). "Mechanical Determinants of Faster Change of Direction Speed Performance in Male Athletes". In: *Journal of Strength and Conditioning Research* 31.3, pp. 696–705. DOI: 10.1519/JSC.0000000000001535.

Draper, J.A. (1985). "The 505 test: A test for agility in the horizontal plane". In: *Australian Journal for Science and Medicine in Sport* 17.1, pp. 15–18.

Edwards, Suzi, Austin, Aaron P., and Bird, Stephen P. (Jan. 2017). "The Role of the Trunk Control in Athletic Performance of a Reactive Change-of-Direction Task". In: *Journal of Strength and Conditioning Research* 31.1, pp. 126–139. DOI: 10.1519/JSC.0000000000001488.

Eiríksson, Eyþór Rúnar (2022). *Understanding Reprojection Error – calib.io*. URL: https://calib.io/blogs/knowledge-base/understanding-reprojection-errors.

Englander, Zoë A, Cutcliffe, Hattie C, Utturkar, Gangadhar M, et al. (2019). "A Comparison of Knee Abduction Angles Measured by a 3D Anatomic Coordinate System Versus Videographic Analysis Implications for Anterior Cruciate Ligament Injury". In: *The Orthopaedic Journal of Sports Medicine* 7.1. DOI: 10.1177/2325967118819831.

Fang, Hao Shu, Li, Jiefeng, Tang, Hongyang, et al. (June 2023). "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.6, pp. 7157–7173. DOI: 10.1109/TPAMI.2022.3222784.

FC Twente (n.d.). *FC Twente | Officiële website van FC Twente Enschede*. URL: https://fctwente.nl/.

Forster, James W D, Uthoff, Aaron M, Rumpf, Michael C, et al. (Oct. 2022). "Pro-agility unpacked: Variability, comparability and diagnostic value". In: *International Journal of Sports Science & Coaching* 17.5, pp. 1225–1240. DOI: 10.1177/17479541211069338.

Geng, Zigang, Sun, Ke, Xiao, Bin, et al. (June 2021). "Bottom-Up Human Pose Estimation Via Disentangled Keypoint Regression". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. IEEE, pp. 14671–14681. DOI: 10.1109/CVPR46437.2021.01444.

Getchell, Bud (1979). *Physical Fitness: A Way of Life. Second Edition*. 2nd. John Wiley & Sons, Inc., One Wiley Drive, Somerset, New Jersey 08873 ($7.95).

Glen, Stephanie (n.d.). *Median Absolute Deviation - Statistics How To*. URL: https://www.statisticshowto.com/median-absolute-deviation/.

Gokeler, A., Welling, W., Benjaminse, A., et al. (Oct. 2017). "A critical analysis of limb symmetry indices of hop tests in athletes after anterior cruciate ligament reconstruction: A case control study". In: *Orthopaedics & Traumatology: Surgery & Research* 103.6, pp. 947–951. DOI: 10.1016/j.otsr.2017.02.015.

GoPro Inc. (n.d.). *GoPro Hero 7 Black | Manual*. URL: https://gopro.com/content/dam/help/hero7-black/manuals/HERO7Black_UM_ENG_REVA.pdf.

Hewit, Jennifer K., Cronin, John B., and Hume, Patria A. (Sept. 2012). "Understanding Change of Direction Performance: A Technical Analysis of a 180° Ground-Based Turn and Sprint Task". In: *International Journal of Sports Science & Coaching* 7.3, pp. 493–501. DOI: 10.1260/1747-9541.7.3.493.

Hidalgo, Gines and Cao, Zhe (2017). *Github | OpenPose Advanced Doc - 3-D Reconstruction Module and Demo*. URL: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/advanced/3d_reconstruction_module.md.

Hidalgo, Gines and Cao, Zhe (n.d.[a]). *Github | CMU-Perceptual-Computing-Lab/openpose/flags.hpp*. URL: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/include/openpose/flags.hpp.

Hidalgo, Gines and Cao, Zhe (n.d.[b]). *Github | OpenPose Doc - Demo*. URL: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/01_demo.md.

Hidalgo, Gines, Raaj, Yaadhav, Idrees, Haroon, et al. (Oct. 2019). "Single-Network Whole-Body Pose Estimation". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Vol. 2019-Octob. IEEE, pp. 6981–6990. DOI: 10.1109/ICCV.2019.00708.

James, Gareth, Witten, Daniela, Hastie, Trevor, et al. (2013). *An Introduction to Statistical Learning*. Vol. 103. Springer Texts in Statistics. New York, NY: Springer New York, pp. 1–29. DOI: 10.1007/978-1-4614-7138-7.

Jin, Sheng, Ma, Xujie, Han, Zhipeng, et al. (2017). "Towards Multi-Person Pose Tracking : Bottom-up and Top-down Methods". In: *International Conference on Computer Vision* 2, pp. 4–7.

Kanko, Robert M., Laende, Elise K., Selbie, W. Scott, et al. (May 2021a). "Inter-session repeatability of markerless motion capture gait kinematics". In: *Journal of Biomechanics* 121, p. 110422. DOI: 10.1016/j.jbiomech.2021.110422.

Kanko, Robert M., Laende, Elise K., Strutzenberger, Gerda, et al. (June 2021b). "Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system". In: *Journal of Biomechanics* 122, p. 110414. DOI: 10.1016/j.jbiomech.2021.110414.

Kassambara, Alboukadel (2019). *Machine Learning Essentials: Practical Guide in R*. 1st. STHDA.

Lea, R.D. and Gerhardt, J.J. (June 1995). "Range-of-motion measurements". In: *The Journal of Bone & Joint Surgery* 77.5, pp. 784–798.

Li, Jianzhu and Valliant, Richard (2011). "Linear regression influence diagnostics for unclustered survey data". In: *Journal of Official Statistics* 27.1, pp. 99–119.

Lovanshi, Mayank and Tiwari, Vivek (Dec. 2022). "Human Pose Estimation: Benchmarking Deep Learning-based Methods". In: *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2022*. IEEE, pp. 1–6. DOI: 10.1109/IATMSI56455.2022.10119324.

Luke, Steven G (Aug. 2017). "Evaluating significance in linear mixed-effects models in R". In: *Behavior Research Methods* 49.4, pp. 1494–1502. DOI: 10.3758/s13428-016-0809-y.

Lundgren, Lina, Tran, Tai T, Farley, Oliver, et al. (2013). "Ankle range of motion among surfing athletes". In: *Journal of Australian Strength and Conditioning* 21.2, pp. 121–124.

Malik, Raza Naseem (2022). "The control of skilled walking: development of novel protocols for assessment of spinal cord injury function and rehabilitation". PhD thesis. University of British Columbia. DOI: 10.14288/1.0417321.

Maloney, Sean J., Richards, Joanna, Jelly, Luke, et al. (Feb. 2019). "Unilateral Stiffness Interventions Augment Vertical Stiffness and Change of Direction Speed". In: *Journal of Strength and Conditioning Research* 33.2, pp. 372–379. DOI: 10.1519/JSC.0000000000002006.

McKay, Brianna D., Miramonti, Amelia A., Gillen, Zachary M., et al. (Apr. 2020). "Normative Reference Values for High School–Aged American Football Players: Proagility Drill and 40-Yard Dash Split Times". In: *Journal of Strength and Conditioning Research* 34.4, pp. 1184–1187. DOI: 10.1519/JSC.0000000000002930.

Meyer, Harald (2019). *Multi Camera Control for GoPro*. URL: https://apps.apple.com/us/app/multi-camera-control-for-gopro/id1449808197.

Movella Inc. (n.d.[a]). *Movella, a Global Leader in Digitization of Movement, to Become Publicly Traded on Nasdaq via Business Combination with Pathfinder Acquisition Corporation*. URL: https://www.movella.com/company/press-room/movella-a-global-leader-in-digitization-of-movement-to-become-publicly-traded-on-nasdaq-via-business-combination-with-pathfinder-acquisition-corporation.

Movella Inc. (n.d.[b]). *MVN Awinda*. URL: https://www.movella.com/products/motion-capture/xsens-mvn-awinda.

Mroz, Sarah, Baddour, Natalie, McGuirk, Connor, et al. (Dec. 2021). "Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose". In: *BioSMART 2021 - Proceedings: 4th International Conference on Bio-Engineering for Smart Technologies*. IEEE, pp. 1–4. DOI: 10.1109/BioSMART54244.2021.9677850.

Mundt, Marion, Born, Zachery, Goldacre, Molly, et al. (2023). "Estimating Ground Reaction Forces from Two-Dimensional Pose Data: A Biomechanics-Based Comparison of AlphaPose, BlazePose, and OpenPose". In: *Sensors* 23.1, pp. 1–15. DOI: 10.3390/s23010078.

Nakagawa, Shinichi and Schielzeth, Holger (Feb. 2013). "A general and simple method for obtaining R
2 from generalized linear mixed-effects models". In: *Methods in Ecology and Evolution* 4.2. Ed. by O'Hara, Robert B., pp. 133–142. DOI: 10.1111/j.2041-210x.2012.00261.x.

Nakano, Nobuyasu, Sakura, Tetsuro, Ueda, Kazuhiro, et al. (May 2020). "Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras". In: *Frontiers in Sports and Active Living* 2.May, pp. 1–9. DOI: 10.3389/fspor.2020.00050.

Nimphius, Sophia, Callaghan, Samuel J., Spiteri, Tania, et al. (Nov. 2016). "Change of Direction Deficit: A More Isolated Measure of Change of Direction Performance Than Total 505 Time". In: *Journal of Strength and Conditioning Research* 30.11, pp. 3024–3032. DOI: 10.1519/JSC.0000000000001421.

OpenCV (n.d.[a]). *Camera calibration With OpenCV*. URL: https://docs.opencv.org/3.4/d4/d94/tutorial_camera_calibration.html.

OpenCV (n.d.[b]). *Detection of ArUco Boards*. URL: https://docs.opencv.org/4.x/db/da9/tutorial_aruco_board_detection.html.

OpenCV (n.d.[c]). *Detection of ArUco Markers*. URL: https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html.

Parks, T.W. and Burrus, C.S. (1987). *Digital Filter Design*. Wiley-Interscience. DOI: 10.5555/27106.

Paul, Darren J, Gabbett, Tim J, and Nassis, George P. (Mar. 2016). "Agility in Team Sports: Testing, Training and Factors Affecting Performance". In: *Sports Medicine* 46.3, pp. 421–442. DOI: 10.1007/s40279-015-0428-2.

Peetsma, Jasper (2023). *OpenPose-COD: Analysis of Change of Direction using OpenPose*. URL: https://github.com/jasperpeetsma/OpenPose-COD.

Philipp, Nicolas M, Crawford, Derek A, Garver, Matthew J, et al. (2021). "Interlimb Asymmetry Thresholds that Negatively Affect Change of Direction Performance in Collegiate American Football Players". In: *International Journal of Exercise Science* 14.4, pp. 606–612.

Postma, Dees B. W., Wieling, Martijn B., Lemmink, Koen A. P. M., et al. (Oct. 2022). "Distance over Time in a Maximal Sprint: Understanding Athletes' Action Boundaries in Sprinting". In: *Ecological Psychology* 34.4, pp. 133–156. DOI: 10.1080/10407413.2022.2120397.

Purcell, Brendan, Channells, Justin, James, Daniel, et al. (Dec. 2005). "Use of accelerometers for detecting foot-ground contact time during running". In: *BioMEMS and Nanotechnology II*. Ed. by Nicolau, Dan V. Vol. 6036, pp. 603–615. DOI: 10.1117/12.638389.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. URL: https://www.r-project.org/.

Rago, Vincenzo, Brito, João, Figueiredo, Pedro, et al. (Feb. 2020). "The Arrowhead Agility Test: Reliability, Minimum Detectable Change, and Practical Applications in Soccer Players". In: *Journal of Strength and Conditioning Research* 34.2, pp. 483–494. DOI: 10.1519/JSC.0000000000002987.

Riazati, Sherveen, McGuirk, Theresa E., Perry, Elliott S., et al. (June 2022). "Absolute Reliability of Gait Parameters Acquired With Markerless Motion Capture in Living Domains". In: *Frontiers in Human Neuroscience* 16, p. 867474. DOI: 10.3389/fnhum.2022.867474.

Roaas, Asbjørn and Andersson, Gunnar B.J. (1982). "Normal range of motion of the hip, knee and ankle joints in Male subjects, 30-40 years of age". In: *Acta Orthopaedica* 53.2, pp. 205–208. DOI: 10.3109/17453678208992202.

Rosenblum, Uri, Lavi, Adi, Fischer, Arielle, et al. (2023). "The Effect of Arm Restriction on Dynamic Stability and Upper Body Responses to Lateral Loss of Balance During Walking: An Observational Study [preprint]". In: *bioRxiv*. DOI: 10.1101/2023.09.11.557158.

Sakurai, Takashi and Okada, Hidetaka (Sept. 2021). "EXAMINATION OF AN APPLICABLE RANGE FOR A MARKERLESS MOTION CAPTURE SYSTEM IN GAIT ANALYSIS". In: *39th International Society of Biomechanics in Sport Conference*. Canberra, Australia: NMU Commons.

Sanchez, Natalia, Schweighofer, Nicolas, and Finley, James M. (2021). "Different Biomechanical Variables Explain Within-Subjects Versus Between-Subjects Variance in Step Length Asymmetry Post-Stroke". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29, pp. 1188–1198. DOI: 10.1109/TNSRE.2021.3090324.

Sandau, Martin, Koblauch, Henrik, Moeslund, Thomas B., et al. (Sept. 2014). "Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane". In: *Medical Engineering & Physics* 36.9, pp. 1168–1175. DOI: 10.1016/j.medengphy.2014.07.007.

Semenick, Doug (Feb. 1990). "Tests And Measurements: The T-test". In: *Strength & Conditioning Journal* 12.1, pp. 36–37.

Sheppard, J.M. and Young, W.B. (Feb. 2006). "Agility literature review: Classifications, training and testing". In: *Journal Of Sports Sciences* 24.9, pp. 919–932. DOI: 10.1080/02640410500457109.

Šimonek, Jaromír, Horička, Pavol, and Hianik, Ján (2016). "Differences in pre-planned agility and reactive agility performance in sport games". In: *Acta Gymnica* 46.2, pp. 68–73.

Stenum, Jan, Rossi, Cristina, and Roemmich, Ryan T. (Apr. 2021). "Two-dimensional video-based analysis of human gait using pose estimation." In: *PLoS computational biology* 17.4. Ed. by Schneidman-Duhovny, Dina, e1008935. DOI: 10.1371/journal.pcbi.1008935.

Tanaka, Ryo, Takimoto, Haruka, Yamasaki, Takahiro, et al. (Apr. 2018). "Validity of time series kinematical data as measured by a markerless motion capture system on a flatland for gait assessment". In: *Journal of Biomechanics* 71, pp. 281–285. DOI: 10.1016/j.jbiomech.2018.01.035.

Teledyne FLIR LLC (n.d.). *Machine Vision Cameras | FLIR Industrial | FLIR Systems*. URL: https://www.flir.eu/browse/industrial/machine-vision-cameras/.

The MathWorks Inc. (n.d.). *MATLAB Camera Calibrator App*. URL: https://nl.mathworks.com/help/vision/ref/cameracalibrator-app.html.

The Mathworks Inc. (n.d.[a]). *3-D locations of undistorted matching points in stereo images - MATLAB triangulate*. URL: https://nl.mathworks.com/help/vision/ref/triangulate.html.

The Mathworks Inc. (n.d.[b]). *Correct point coordinates for lens distortion - MATLAB undistortPoints*. URL: https://nl.mathworks.com/help/vision/ref/undistortpoints.html.

The Mathworks Inc. (n.d.[c]). *Decode JSON-formatted text - MATLAB jsondecode*. URL: https://nl.mathworks.com/help/matlab/ref/jsondecode.html.

The Mathworks Inc. (n.d.[d]). *Detect and remove outliers in data - MATLAB rmoutliers*. URL: https://nl.mathworks.com/help/matlab/ref/rmoutliers.html.

The Mathworks Inc. (n.d.[e]). *Detect and replace outliers in data - MATLAB filloutliers*. URL: https://nl.mathworks.com/help/matlab/ref/filloutliers.html.

The Mathworks Inc. (n.d.[f]). *Numerical gradient - MATLAB gradient*. URL: https://nl.mathworks.com/help/matlab/ref/gradient.html.

Theia Markerless Inc. (n.d.). *Theia Markerless - Markerless Motion Capture Redefined*. URL: https://www.theiamarkerless.ca/.

Thomas, Christopher, Dos'Santos, Thomas, Comfort, Paul, et al. (May 2020). "Effect of Asymmetry on Biomechanical Characteristics During 180° Change of Direction". In: *Journal of Strength and Conditioning Research* 34.5, pp. 1297–1306. DOI: 10.1519/JSC.0000000000003553.

Trackwired (n.d.). *YouTube | BOUNDING: Bounding Routine Alternate Leg Bounds*. URL: https://www.youtube.com/watch?v=b3124L0KK3Q&ab_channel=Trackwired.

Washabaugh, Edward P., Shanmugam, Thanikai Adhithiyan, Ranganathan, Rajiv, et al. (Sept. 2022). "Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics". In: *Gait & Posture* 97, pp. 188–195. DOI: 10.1016/j.gaitpost.2022.08.008.

Webering, Fritz, Blume, Holger, and Allaham, Issam (June 2021). "Markerless camera-based vertical jump height measurement using OpenPose". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 3863–3869. DOI: 10.1109/CVPRW53098.2021.00428.

Welch, Neil, Richter, Chris, Franklyn-Miller, Andy, et al. (Jan. 2019). "Principal Component Analysis of the Biomechanical Factors Associated With Performance During Cutting". In: *Journal of Strength and Conditioning Research* Publish Ah.6. DOI: 10.1519/JSC.0000000000003022.

Welling, Wouter and Frik, Laurens (Dec. 2021). "On-Field Tests for Patients After Anterior Cruciate Ligament Reconstruction: A Scoping Review". In: *Orthopaedic Journal of Sports Medicine* 10.1. DOI: 10.1177/23259671211055481.

Wieling, Martijn (Sept. 2018). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English". In: *Journal of Phonetics* 70, pp. 86–116. DOI: 10.1016/J.WOCN.2018.03.002.

Wilm, Jakob (n.d.). *Calibration Patterns Explained*. URL: https://calib.io/blogs/knowledge-base/calibration-patterns-explained.

Winter, Bodo (Aug. 2013). "Linear models and linear mixed effects models in R with linguistic applications". In: *arXiv preprint arXiv:1308.5499*, pp. 1–42.

Xu, Yufei, Zhang, Jing, Zhang, Qiming, et al. (2022). "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation". In: *Advances in Neural Information Processing Systems* 35.NeurIPS, pp. 1–14.

Young, W.B., Dawson, B., and Henry, G.J (2015). "Agility and Change-of-Direction Speed are Independent Skills: Implications for Training for Agility in Invasion Sports". In: *International Journal of Sports Science & Coaching* 10.1, pp. 159–169.

Young, W.B., Dos'Santos, T., Harper, D., et al. (2022). "Agility in Invasion Sports: Position Stand of the IUSCA". In: *International Journal of Strength and Conditioning* 2.1, pp. 1–25. DOI: 10.47206/ijsc.v2i1.126.

Zago, Matteo, Luzzago, Matteo, Marangoni, Tommaso, et al. (Mar. 2020). "3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision". In: *Frontiers in Bioengineering and Biotechnology* 8, p. 181. DOI: 10.3389/fbioe.2020.00181.

Zatsiorsky, V., Seluyanov, V., and Chugunova, L. (1990). "In vivo body segment inertial parameters determination using a gamma-scanner method". In: *Biomechanics of Human Movement: Applications in Rehabilitation, Sports and Ergonomics*. Ed. by Berme, N. and Cappozzo, A. Vol. 7. Worthington, Ohio: Bertec Corporation, pp. 186–202.

Zhang, Zhengyou (Feb. 2012). "Microsoft Kinect Sensor and Its Effect". In: *IEEE Multimedia* 19.2, pp. 4–10. DOI: 10.1109/MMUL.2012.24.

Appendix A

Information Letter

Beste spelers en ouders/verzorgers,

Recent hebben we u geïnformeerd over de testdagen die gepland staan voor maandag en dinsdag, 4 en 5 juli. De testdagen zijn voornamelijk bedoeld om inzicht te krijgen in de medische en fysieke staat van onze spelers, dit ter optimalisatie van prestatie en het voorkomen van blessures. Met de inzichten die we opdoen, kunnen we onze spelers gerichte en individuele begeleiding bieden.

Tijdens de testdagen zullen er, naast onze eigen medische staff, fysieke staff, en wetenschappelijke staff, tevens onderzoekers van de Universiteit Twente aanwezig zijn. Zij zullen onderzoek doen naar wendbaarheid in voetbal. Op verzoek van de Universiteit Twente verschaffen we u bij dezen met aanvullende informatie aangaande hun onderzoek.

Wat houdt het onderzoek van de Universiteit Twente in?
Wendbaarheid wordt meestal gemeten door gevalideerde sprint- en wendbaarheidstesten. De snelheid waarmee een speler een bepaalde afstand of parcours aflegt wordt gezien als een goede indicatie voor de wendbaarheid van deze speler. Tijd en snelheid geven trainers en coaches echter maar weinig inzicht in de redenen dat een speler wendbaar is. Sommige spelers hebben een grote wendbaarheid omdat ze beschikken over een groot explosief vermogen, terwijl andere spelers het juist moeten hebben van hun uitstekende coördinatie. Onderzoekers van de Universiteit Twente onderzoeken of ze de motorische kwaliteiten van spelers kunnen linken aan wendbaarheid. De resultaten van dit onderzoek kunnen trainers en coaches helpen om gerichtere trainingsinterventies te ontwikkelen ten behoeve van wendbaarheid.

Welke data wordt er verzameld voor het onderzoek?
Voor het onderzoek naar wendbaarheid zal er data gedeeld worden met de Universiteit Twente en zal er data verzameld worden door de Universiteit Twente. Gegevens uit prestatietests, zijnde de '5-0-5 test' en de 'arrow-head test'; gegevens uit sporttrackers, zijnde acceleratie en hartslagdata; en antropometrische gegevens (zoals lengte en gewicht) en persoonsgegevens (leeftijd) zullen gedeeld worden met de Universiteit Twente. In aanvulling zullen onderzoekers van de Universiteit Twente videodata opnemen van spelers voor de '5-0-5 test' en de 'arrow-head test'.

Hoe verwerkt de Universiteit Twente deze data?
Zowel de data die gedeeld wordt met de Universiteit Twente alswel de data die verzameld wordt door de Universiteit Twente wordt geanonimiseerd. Voor de videodata betekent dit dat de gezichten van de spelers onherkenbaar worden gemaakt. De onderzoeksgegevens worden, conform de richtlijnen van de VNSU, minimaal 10 jaar bewaard. De data, waaronder de geanonimiseerde videodata, is enkel toegankelijk voor de mensen die betrokken zijn bij het onderzoeksprogramma. Een lijst met namen van mensen die toegang hebben tot het materiaal is beschikbaar en kan worden opgevraagd via Youri Geurkink, coördinator Fysiek en Wetenschap van FC Twente/Heracles Academie. De materialen zullen op geen enkele wijze publiek beschikbaar gemaakt worden of gebruikt worden voor promotiedoeleinden. De geanonimiseerde data kan enkel gebruikt worden voor onderzoeksdoeleinden.

Kan ik afzien van deelname aan het onderzoek van de Universiteit Twente?
Spelers zijn niet verplicht om deel te nemen aan het onderzoek. Mochten u af willen zien van deelname aan het onderzoek van de Universiteit Twente, dan kunt u dit aangeven bij Youri Geurkink. Deze beslissing zal niet van invloed zijn op de deelname van de speler tijdens de testdagen. Mocht

u na afloop van de testdagen alsnog af willen zien van deelname, dan kunt u dat binnen 5 werkdagen doorgeven aan Youri Geurkink. De opgenomen videodata zal dan verwijderd worden.

Met wie kan ik contact opnemen voor aanvullende vragen?
Voor vragen omtrent de testdagen kunt u contact opnemen met Youri Geurkink, Coördinator Fysiek en Wetenschap, FC Twente/Heracles Academie (y.geurkink@fctwenteheraclesacademie.nl)

Voor vragen omtrent het onderzoek naar wendbaarheid kunt u contact opnemen met de hoofdonderzoeker op dit project, Dees Postma, Assistant Professor, Universiteit Twente (d.b.w.postma@utwente.nl).

Wilt u graag een onafhankelijk advies over meedoen aan dit onderzoek, of een klacht indienen? Dan kunt u terecht bij de secretaris van de Ethische Commissie (ethicscommittee-cis@utwente.nl). De commissie bestaat uit onafhankelijk deskundigen van de universiteit en is beschikbaar voor vragen en klachten rondom het onderzoek.

Appendix B

Pilot Test Documents

B.1 Participant Information

I would like to invite you to take part in a research study for my graduation project. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read needs clarification or if you would like more information. Take time to decide whether or not to take part.

Who I Am And What This Study Is About

My name is Jasper Peetsma, and currently, I am doing my graduation project for my master's in Interaction Technology at the University of Twente. During this project, I focus on defining new performance indicators for agility, quantifying these, and measuring them using open-source video-based human pose detection software. The data can be used by trainers and sports scientists to better understand the difference in performance between players and to track whether specific performance indicators improve or decline over time. The research activity you are invited to participate in aims to collect movement data for further analysis as part of the project.

What Will Taking Part Involve?

The activity involves performing so-called change-of-direction tests wearing a motion capture (MoCap) suit while being video-recorded for biomechanical analysis. The data will be used to study the use of video data to quantify agility. The session will take an hour at maximum. The location of the session is Pro-F, Kotkampweg 65.

Why Have You Been Invited To Take Part?

You have been invited to this research activity because, as a professional football player, you are part of the target group that the system is designed for. Therefore you are considered a potential participant who can provide a valuable and realistic performance in the change-of-direction test.

Do You Have To Take Part?

Your participation in this research activity is entirely voluntary. You have the right to refuse participation, refuse any question or assignment, and withdraw at any time without any consequence whatsoever.

What Are The Possible Risks And Benefits Of Taking Part?

Participation in this research activity provides valuable data that can be used to better understand agility. You are asked to perform the test the best you can. However, as each performance is measured using two methods that will be compared, any performance is helpful for research. All input is good input.

Just like in any other physical exercise, there is a risk of injury. However, this risk is equal to any other common change-of-direction or agility test. Should anything happen during one of the tests, you are free to stop participating at any time without any consequence whatsoever.

Will Taking Part Be Confidential?

Anything you say or do for the full duration of the session will remain confidential. The video footage and the MoCap data are retained as part of the research process. However, they will not be shared with anyone outside the researchers and employees involved in the graduation project. After analysis, the data will be anonymised and might be used for publication. Any frames from the videos taken during the test may be used in the thesis. Any personally identifiable information will be anonymised by, for example, blurring your face and any other aspects that might be enough to recognise you as a person. This way, no data used in the report can in any way be connected to you as a person, except by me as the researcher due to being present at the session.

How Will Information You Provide Be Recorded, Stored And Protected?

As stated earlier, the information you provide during the session by performing the tests will be recorded through video and full body motion capture. The research data will be stored securely on a password-protected computer and backed up in a password-protected online storage account. Access to the research data will be limited to only me. The video recordings will be deleted once the graduation project is finished. However, the anonymised research data used in my thesis will be retained by UT staff involved in the project. You have the right to request rectification or erasure of personal data until one week after participation. You have the right to request access to personal data at any time.

What Will Happen To The Results Of The Study?

The session results will be used for the writing of my thesis and as input for the following stages of my research design. This includes MoCap data, the video recording of the test session, and the findings based on your input.

Whom Should You Contact For Further Information?

For any further information, you can contact the researcher, Jasper Peetsma, by sending an email to j.j.peetsma@student.utwente.nl.

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please get in touch with Dees Postma, my supervisor of the graduation project, through the email d.b.w.postma@utwente.nl.

Additionally, you can contact the Secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente through ethicscommittee-cis@utwente.nl.

Thank You For Considering Participating In This Research Study.

Your experiences and opinions are indispensable in this field of study, and completing this study wouldn't be possible without your help.

B.2 Consent Form

Please tick the appropriate boxes.	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 13-04-2023, or it has been read to me. I have been able to ask questions about the study, which have been answered to my satisfaction.	○	○
I consent voluntarily to participate in this study and understand that I can refuse to answer questions and withdraw from the study at any time without having to give a reason.	○	○
I understand that participating in the study involves performing a video-recorded agility test while wearing a motion capture suit.	○	○
Risks associated with participating in the study		
I understand that participating in the study involves the risk of physical discomfort or injury, but this risk is no greater than in any other standard agility test.	○	○
Use of the information in the study		
I understand that the information I provide will be used for the writing of the researcher's thesis for the Master program Interaction Technology.	○	○
I understand that personal information collected about me that can identify me, such as my name, will only be shared with the researchers and employees involved in the graduation project.	○	○
I agree that the information I provide can be used in research outputs as long as they are anonymised.	○	○
Consent to be video-recorded and motion-captured		
I agree to be video-recorded during the session, and I understand that the video-recorded data will be used for analysis, anonymised, and retained as part of the research process.	○	○
I agree to be full-body motion-captured during the session, and I understand that the motion-capture data will be used for analysis, anonymised, and retained as part of the research process.	○	○
Future use and reuse of the information by others		
I give permission for the anonymised data I provide to be archived on the researcher's computer to be used for future research and learning.	○	○
Signatures		

Name of participant	Signature	Date
I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.		

Name of researcher	Signature	Date

Appendix C

Dataset

Variable	Category	Unit
PID	Factor	range [0,50]
age	Demographic	years
weight	Anthropometric	kg
height		cm
sittingHeight		cm
CODTest	Factor	'arrowhead'
direction	Factor	'L' / 'R'
CODTestTime	Test Results	*s*
sprint10m1st		*s*
sprint10m2nd		*s*
sprint10mAvg		*s*
sprint30m1st		*s*
sprint30m2nd		*s*
sprint30mAvg		*s*
AFCT505Left		*s*
AFCT505Right		*s*
AFCT505Avg		*s*
AFCTFullLeft		*s*
AFCTFullRight		*s*
AFCTFullAvg		*s*
Left505		*s*
Right505		*s*
Avg505		*s*
arrowheadLeft		*s*
arrowheadRight		*s*
arrowheadAvg		*s*
endurance1000m		*s*
endurance400m		*s*
meanLeftStanceTime	Stance Time	*s*
meanRightStanceTime		*s*
meanBilateralStanceTime		*s*
minLeftStanceTime		*s*
minRightStanceTime		*s*
minBilateralStanceTime		*s*
maxLeftStanceTime		*s*
maxRightStanceTime		*s*
maxBilateralStanceTime		*s*
totalLeftStanceTime		*s*
totalRightStanceTime		*s*
meanL2RStepTime	Step Time	*s*
meanR2LStepTime		*s*
meanBilateralStepTime		*s*
minL2RStepTime		*s*
minR2LStepTime		*s*
minBilateralStepTime		*s*
maxL2RStepTime		*s*
maxR2LStepTime		*s*
maxBilateralStepTime		*s*

Variable (continued)	Category (continued)	Unit (continued)
meanLeftSwingTime	Swing Time	*s*
meanRightSwingTime		*s*
meanBilateralSwingTime		*s*
minLeftSwingTime		*s*
minRightSwingTime		*s*
minBilateralSwingTime		*s*
maxLeftSwingTime		*s*
maxRightSwingTime		*s*
maxBilateralSwingTime		*s*
totalDoubleSupportTime	Double Support Time	*s*
doubleSupportPercentage		%
meanRightStepLength	Step Length	*m*
meanLeftStepLength		*m*
meanBilateralStepLength		*m*
minRightStepLength		*m*
minLeftStepLength		*m*
minBilateralStepLength		*m*
maxRightStepLength		*m*
maxLeftStepLength		*m*
maxBilateralStepLength		*m*
runningGaitSpeed	Running Gait	m/s
kneeLowRangeDisbalance	Disbalance	range [0,1]
kneeMidRangeDisbalance		range [0,1]
kneeHighRangeDisbalance		range [0,1]
hipLowRangeDisbalance		range [0,1]
hipMidRangeDisbalance		range [0,1]
hipHighRangeDisbalance		range [0,1]
ankleLowRangeDisbalance		range [0,1]
ankleMidRangeDisbalance		range [0,1]
ankleHighRangeDisbalance		range [0,1]
kneeRangeOfMotionLeft	Range of Motion	degrees
kneeRangeOfMotionRight		degrees
kneeRangeOfMotionDiff		degrees
hipRangeOfMotionLeft		degrees
hipRangeOfMotionRight		degrees
hipRangeOfMotionDiff		degrees
ankleRangeOfMotionLeft		degrees
ankleRangeOfMotionRight		degrees
ankleRangeOfMotionDiff		degrees
meanLeftTakeoffDistance	Takeoff Distance	*m*
meanRightTakeoffDistance		*m*
meanBilateralTakeoffDistance		*m*
maxLeftTakeoffDistance		*m*
maxRightTakeoffDistance		*m*
maxBilateralTakeoffDistance		*m*
meanSpeed	Speed	m/s
maxSpeed		m/s
meanAcceleration	Acceleration	m/s^2
minAcceleration		m/s^2
maxAcceleration		m/s^2
meanPosAcceleration		m/s^2
meanNegAcceleration		m/s^2