KTH ROYAL INSTITUTE
OF TECHNOLOGY

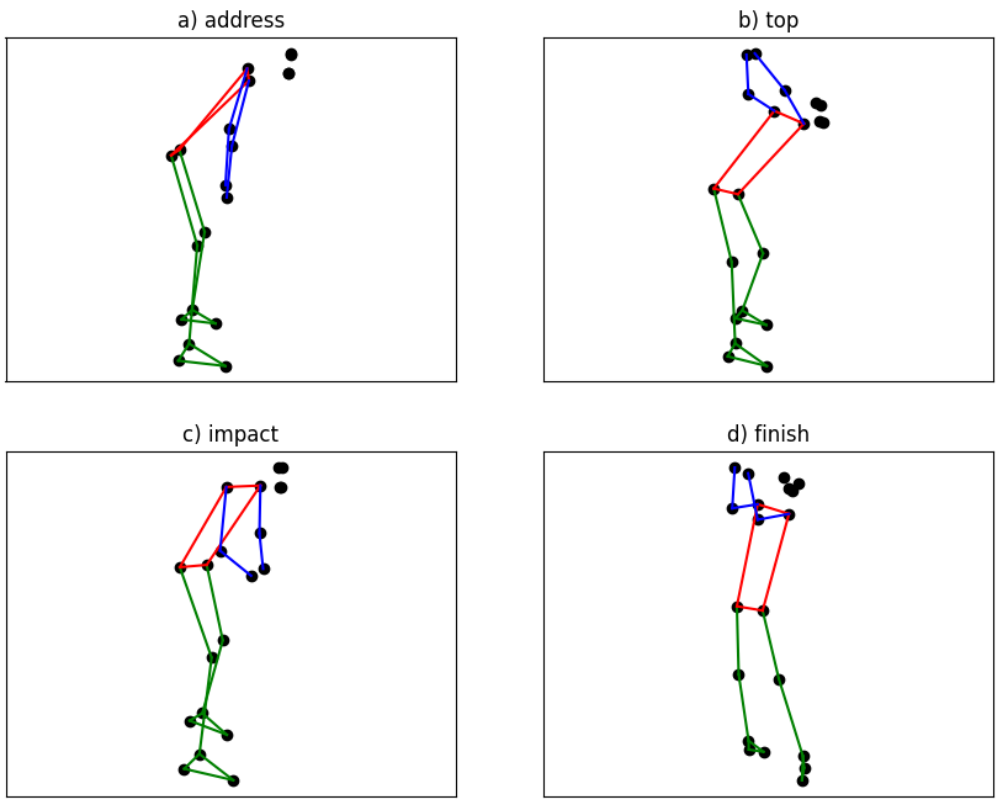
Degree Project in Sports Technology

Second Cycle 30 ECTS

Feasibility of Mobile Phone-Based 2D Human Pose Estimation for Golf

An analysis of the golf swing focusing on selected joint angles

ELISA PERINI



Stockholm, Sweden 2023

Feasibility of Mobile Phone-Based 2D Human Pose Estimation for Golf

An analysis of the golf swing focusing on selected joint angles

ELISA PERINI

Master's programme in Sports Technology
School of Engineering Sciences in Chemistry, Biotechnology and Health
KTH Royal Institute of Technology

TRITA:	CBH-GRU-2023:057
Date:	14th May 2023
Supervisors:	Felix Eder, Andrei Dubavets
Reviewer:	Jonas Willén
Examiner:	Sebastiaan Meijer
Host company:	TopTracer
Swedish title:	Lämpligheten av mobiltelefonbaserad 2D mänsklig poseuppskattning i golf: en analys av golfsvingar med fokus på utvalda ledvinklar

Abstract

Golf is a sport where the correct technical execution is important for performance and injury prevention. The existing feedback systems are often cumbersome and not readily available to recreational players. To address this issue, this thesis explores the potential of using 2D Human Pose Estimation as a mobile phone-based swing analysis tool.
The developed system allows to identify three events in the swing movement (toe-up, top and impact) and to measure specific angles during these events by using an algorithmic approach. The system focuses on quantifying the knee flexion and primary spine angle during the address, and lateral bending at the top of the swing. By using only the wrist coordinates in the vertical direction, the developed system identified 37% of investigated events, independently of whether the swing was filmed in the frontal of sagittal frame. Within five frames, 95% of the events were correctly identified. Using additional joint coordinates and the event data obtained by the above-mentioned event identification algorithm, the knee flexion at address was correctly assessed in 66% of the cases, with a mean absolute error of 3.7°. The mean absolute error of the primary spine angle measurement at address was of 10.5°. The lateral bending angle was correctly identified in 87% of the videos.
This system highlights the potential of using 2D Human Pose Estimation for swing analysis. This thesis primarily focused on exploring the feasibility of the approach and further research is needed to expand the system and improve its accuracy. This work serves as a foundation, providing valuable insights for future advancements in the field of 2D Human Pose Estimation-based swing analysis.

Keywords

Golf, Human Pose Estimation, Sports Analytics, Computer Vision

Sammanfattning

Golf är en sport där korrekt tekniskt utförande är avgörande för prestation och skadeförebyggelse. Feedbacksystem som finns är ofta besvärliga och inte lättillgängliga för fritidsspelare. För att åtgärda detta problem undersöker detta examensarbete potentialen att använda 2D mänsklig poseuppskattning som mobiltelefonsbaserat svinganalysverktyg.
Det utvecklade systemet gör det möjligt att identifiera tre händelser i svingen (*toe-up*, *top* och *impact*) och att mäta specifika vinklar under dessa händelser genom en algoritmisk metod. Systemet fokuserar på att kvantifiera knäböjningen och primära ryggradsvinkeln under uppställningen, och laterala böjningen vid svingtoppen. Genom att endast använda handledskoordinater i vertikal riktning identifierade det utvecklade systemet 37% av de undersökta händelserna oavsett om svingen filmades från frontal- eller medianplanet. Inom fem bildrutor identifierades 95% av händelserna korrekt. Genom att använda ytterligare ledkoordinater och händelsedata som erhållits genom den tidigare nämnda algoritmen för händelseidentifiering, bedömdes knäböjningen vid uppställningen vara korrekt i 66% av fallen med en medelabsolutfel på 3.7°. Medelabsolutfelet för mätningen av primär ryggradsvinkel vid uppställningen var 10.5°. Laterala böjningen identifierades korrekt i 87% av tillfällena.
Detta system belyser potentialen i 2D mänsklig poseuppskattning för svinganalys. Detta examensarbete fokuserade främst på att utforska tillvägagångssättets genomförbarhet och ytterligare forskning behövs för att utveckla systemet och förbättra dess noggrannhet. Detta arbete är grundläggande och ger värdefulla insikter för framtida forskning inom området för svinganalys baserad på 2D mänsklig poseuppskattning.

Nyckelord

Golf, Mänsklig Poseuppskattning, Sportanalys, Datorseende

Table of contents

Acronyms

API Application Programming Interface

HPE Human Pose Estimation

IMU Inertial Movement Unit

PCA Percentage of Correct Angles

PCE Percentage of Correct Events

PCK Percentage of Correct Keypoints

List of figures

List of Tables

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors Andrei and Felix. Their knowledge, expertise, and availability were very helpful and inspiring, and our Monday morning syncs were crucial for both the structuring and successful completion of this thesis. I am also very grateful to my reviewer Jonas for his insightful review and for his willingness to answer any of my questions. Thanks should also go to my supervision group peers and all other people who took time to read through my writing for their useful comments and feedback.

I had the pleasure of working at Toptracer and in the TRCA team. The warm welcome, and the freedom to explore and try out cool stuff that I was given were very appreciated. I will definitely miss Wednesday morning breakfasts. This thesis could not have been possible without all the people at the Toptracer office who agreed to let me film them, thank you for being great study participants, and for the insights into golf technique.

Big thanks should go to my parents for encouraging my curiosity since I was kid, and to my friends for just generally being awesome. Thanks to the sports technology people for the friendship and for being the absolute best study buddies one could wish for. This degree would not have been the same without them and I wish them all the best in their future endeavours. Finally, thank you Mathias for the love, kindness and patience, and for showing me that life after graduation is not as scary as I always thought it would be.

Stockholm, 14th May 2023
Elisa Perini

Chapter 1

Introduction

Golf as a sport has grown in popularity over the past years and the amount of golf players has risen consistently throughout the last years in the US [1]. This rise in popularity coincides with the sport being reintroduced to the Olympic Games in 2016. The main goal of golf is simple: the player is supposed to hit a ball into a hole in a minimal number of strokes. The first shot generally aims at moving the ball far away and is called a driver swing. For a good driver swing, the golfer needs to hit the ball with the club in a way that the ball achieves maximal distance in the desired direction. To do this, the clubhead should be travelling at maximal speed at the impact moment between the club and the ball [2].

1.1 Problem

The correct and constant execution of the swing is important both in terms of injury prevention and performance increase [3]. While golf is not generally considered an injury-prone sport, overuse injuries are relatively common amongst recreational and professional athletes [4]. Improving the swing is mostly based on observation, either of better players or from a coach giving feedback. Even though swing analysis tools exist, these are not always available, especially amongst recreational players that would benefit the most from technique correction. Providing golfers with an easy-to-use, portable feedback system that they could simply use on their phones, could give them good insights about their technique. This system would not only be beneficial to golfers without a coach but could also be used by coaches as a tool to help obtain more quantitative data about their athletes' swing. The benefits of such a system are interesting both in terms of performance increase, and in terms of injury prevention, as poor technique has been linked with a higher injury risk [5, 6]. Additionally, the use of sports analytics is nowadays widespread in the sports field, and has been associated with improved training practices, and activity enjoyment both for the athlete and the spectator [7]. For instance, such a golf swing analysis system could detect lead side lateral bending, a type of technical mistake, which happens when the golfer leans too much towards the target. For a golfer, one of the best and easiest ways to ensure the correct form of the swing is to film themselves. However, without any additional information, the obtained video might not be that useful for improving or correcting the swing movement. This is particularly true for a recreational golfer who does not have the expert eye of a coach available. Additionally, a coach might deplore the lack of quantitative information, such as joint angles, that makes it difficult to compare swings to track the progress of their athlete. Current solutions to track the golfer's pose and analyse the golf swing are mostly equipment heavy, needing multiple cameras, or depth cameras, which results in systems that are only installed in specific locations or systems not yet broadly available. One solution for this is Human Pose Estimation, a computer vision technique, which provides a reliable marker-less technique to track the joints of the human body in 2D or 3D in real-time from visual information, such as a videos or images. The obtained skeleton model of the human can be used to enhance the feedback from the execution of the movement. Because of its potential, its use in the sports and health field has drastically increased in the last ten years [8]. In sports and physical activity, this real-time tracking can provide insightful feedback about the body's position by providing, for instance, quantitative information about joint alignment. With the drastic improvement of computational power in the last ten years, human pose detection can now be run on a mobile phone. This allows for a very portable feedback system, that can be carried in the golfer's pocket.

1.2 Purpose and hypothesis

The purpose of this thesis is to investigate whether 2D Human Pose Estimation, that can run on any recent phone, could be used to provide some quantitative information about the swing, such as joint angles, and qualitative information, such as identifying some commonly made technical mistakes. This could be beneficial for beginners that do not have the time nor the knowledge required for identifying their technique errors [9]. The project started with the hypothesis that while 2D Human Pose Estimation has some limits due to missing depth information which will prevent the analysis of some aspects of the swing, a lot of information, such as joint angles or certain technical errors may be obtained from a simple video recorded on a phone.

1.3 Goals

To answer the hypothesis, five goals were identified.

- Identify features and technical errors in the golf swing that can be identified using 2D Human Pose Estimation.
- Create a prototype of the application allowing to obtain the joint coordinates from a swing video, preferably in real-time.
- Identify the phases of the swing in the video to give some context to the coordinate data. To do this, two approaches are possible: one using a machine learning model or one using an algorithmic approach based on typical patterns observed in the joint coordinate data at various moments throughout the swing. The more suitable one for the purpose of this thesis should be chosen.
- Create an algorithm that can identify the chosen features or technical mistakes.
- Obtain the data, either by collecting it or using an existing dataset, to validate the developed system.

The ideal final objective, beyond the scope of this thesis, would be to turn the prototype into an application that can analyse the golf swing and detect eventual mistakes in it, and give feedback and provide data to the user.

1.4 Scope and limitations

To limit the scope of the thesis, only 2D Human Pose Estimation was considered. While some of the used methods may provide 3D information as well, this was not be used. In the same way, only technical aspects that could be detected with the two-dimensional limit were be considered. The project was limited in time and, in consequence, did not aim at identifying and listing all the features from the swing that can be detected using 2D Human Pose Estimation. Discriminating between a technically correct and incorrect swing was also out of the scope. Finally, the thesis neither aimed at making a difference between the types of swing, nor at identifying them. Throughout the thesis, when talking about a swing, a driver swing is meant. Similarly, all analysed swings were driver swings.

Chapter 2

Background

This chapter goes over the background knowledge required to understand the proposed swing analysis tool. It introduces the technical aspects of a golf swing, how Human Pose Estimation (HPE) is currently used in sports and goes over the existing solutions for swing analysis. When it comes to golf technique, it aims at providing some understanding of golf basics to the reader without previous knowledge about golf.

2.1 The phases of the golf swing

To allow for a common ground of reference when describing the golf swing, laterality and phases of the swing can be described. The swing is a complex asymmetric movement, and two sides can be defined: the lead side, also called dominant side which is the side closer to the target which is usually the hole, and the trail side or non-dominant side, furthest from the target. For a right-handed golfer, the left side is the lead side, and the right side the trail side, and vice versa for a left-handed golfer. Research has typically categorized the swing movement into four phases. These include address (A in Figure 2.1), backswing (B45 to B225, where the number after the letter indicates the angle in degrees between the club and the vertical hip-ball line), downswing (D225 to D45) and follow-through (not shown in Figure 2.1), where the backswing and downswing are separated by the top of the swing, also called transition (TP to TC), and the impact (I) separates the downswing and the follow-through [10]. Most of the research in golf uses some variant of this classification, sometimes defining additional mid-phases.

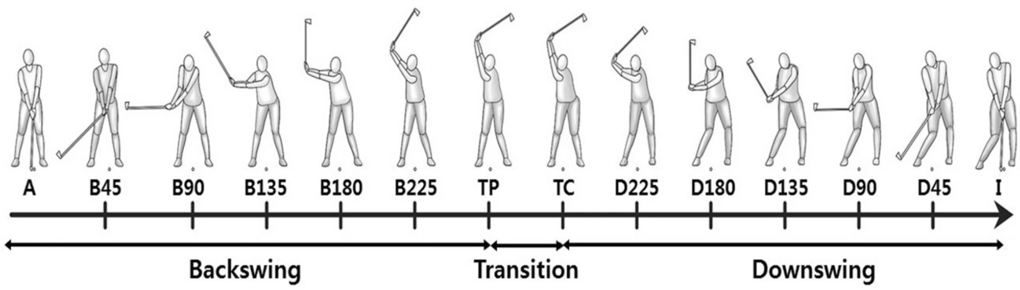


Figure 2.1: Beginning of the golf swing until the impact, showing three of the phases. Address (A), backswing (B) and downswing (D) illustrated with the angle of the club, where 0° is the angle of the club at impact. Image: [11] licensed under CC BY 4.0

2.1.1 Address

The swing starts with the address, also sometimes called set-up: this is when the golfer prepares for the swing. The golfer ensures a good and stable position, correct positioning of the feet, and a firm grip to the club, which is pointing towards the ground. The feet should generally be pointing at 90° from the target, at which the golfer wants to aim, but there can be some individual variation depending on the mobility of the hips [12]. The aim is to be aligned with the target, establish balance and provide an effective grip on the club [13]. The shoulders should be relaxed and the knees should be flexed to 20-25 degrees [13, 14]. Murakami et al. [15], who performed an X-ray analysis of the knees during a golf swing, found the knee angles to be 18±12 degrees for the leading side and 17±9 degrees for the trail side, the study did not however specify whether the participants were experienced golfers. The torso should be bent forward: the primary spine angle, which is the angle between the hip-knee line and the torso, should be of about 45 degrees [13, 14]. Similarly, Kim et al. [16] found the flexion of the hip to be 40 degrees amongst professional male golfers with normal hip mobility, lower for golfers with limited hip mobility. The flexion of the knees allows to bring the club-head to the ground without any excessive flexion of the hips [17]. Figure 2.2 illustrates how the knee flexion and the primary spine angle are measured on a skeleton model of a golfer during the address.

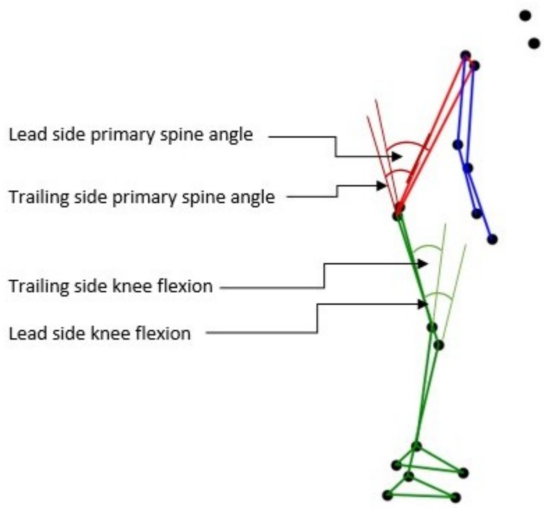


Figure 2.2: Schema of the measured angles

2.1.2 Backswing

The address is followed by the backswing when the golfer brings the club upwards on the trail side and eventually over their head. This phase prepares the downswing and supplies the player with the energy required for the downswing and impact with the ball. The shoulders should follow the movement and turn towards the trail side.

2.1.3 Transition

In between the backswing and the downswing, there is the top of the swing or transition. This is where the club-head reaches its most lateral position, before changing direction [2]. The shoulder or upper torso rotation angle is about 78-102 degrees, and the hip or pelvic rotation angle is about 47-55 degrees [13]. Zhang et al. [3] found the transition time to last on average 0.021 seconds amongst a group of twenty-two experienced male golfers. The stop in the movement during the transition time allows to correctly load for the downswing. An important factor linked to performance is the X-factor, also called torso-pelvic separation, which is the dissociation between the scapular and pelvic girdles. As shown in Figure 2.3, it is the angle of the crossing of the line through the shoulders and the line through the pelvis when the player is viewed from the top in the transverse plane. The X-factor angle is generally computed from the transverse plane or the swing plane, which is the plane formed by the clubhead's swinging [10]. A longer transition time maximises the X-factor, which correlates with a faster ball exit and thus a more powerful swing [3]. This can also be seen in an increase in the elastic potential energy in the trunk muscles, which leads to a better performance [10].

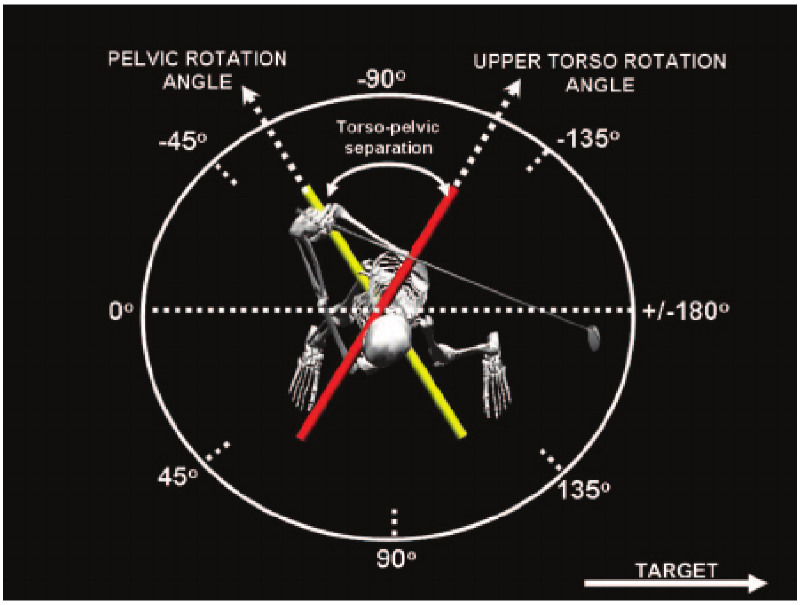


Figure 2.3: Top view of the X-Factor or torso-pelvic separation. Image: [18]

2.1.4 Downswing

During the downswing, the golf club is accelerated down towards the ball. This movement produces a significant amount of kinetic energy which is transferred to the ball on impact, at the end of the phase. In the downswing it is important to start the movement with the lower body and the core, the arms just follow the movement, so that the golfer can put more power into their swing. The torso rotates, starting with the hips and finishing with the shoulders. By beginning the downswing with the rotation of the pelvis, the player can increase its X-factor [10]. The largest value for the X-factor should not be reached during the transition, but at the very beginning of the phase [19], as the hips should start rotating before the shoulders.

2.1.5 Impact

The impact separates the downswing from the follow-through. It is generally considered as one of the most important moments in the swing, because of its importance for aiming. During the impact, the head should be in line with the spine, and not bend forwards or backwards. The primary spine angle should be similar to the one during the address [13, 14].

2.1.6 Follow-through

From the impact starts the follow-through, where the movement decelerates, remaining kinetic energy dissipates, and the club goes back upwards and towards the lead side. Throughout the swing the weight distribution on the legs shifts, starting at an approximately equal distribution during the address and ending with about 60% on the lead side. This transfer of weight is more marked in more skilled players [20, 21, 22]. During the swing the back should be held in a neutral position, without slouching the shoulders.

2.1.7 Phases in practice

In studies that use more phases, mid-phases are defined. These mid-phases are typically defined with, for instance, the club's shaft being parallel with the ground during the downswing or the lead arm being parallel with the ground [23]. In practice, outside of research and literature, professional and recreational players classify the swing in more or less phases, depending on the preferences of the player or the coach [12]. One of the systems used is the P classification system, which divides the swing in ten phases, and has some overlap with the systems used in research as well. The identification of the phases has been done in many different ways, including by qualitatively assessing through videos [22], using muscle activation [24], or machine learning methods. For instance, SwingNet, which is a lightweight deep neural network [25], identifies eight key-points in the swing from a video: address, toe-off, mid-backswing, top of the swing, mid-downswing, impact, mid-follow-through and finish.

2.2 Common mistakes in the swing

This section consists in a short overview of some common technical mistakes made in golf, as well as some that could potentially be identified using 2D human pose detection.

2.2.1 Position at the address and at impact

The address position can be a cause of multiple errors.

- Beginners might not keep their back in a neutral position and arch their back too much [12].
- The primary spine angle can be higher, and this will result in an excessive thoracic flexion. This moves the centre of gravity forwards, limits the spine rotation during the backswing, and eventually leads to a less powerful swing [14].
- As stated in Section 2.1.1, since the knee flexion has an influence on the primary spine angle, more extended or flexed knees can cause excessive thoracic flexion as well.

Similarly, mistakes can be made during the impact.

- As during the address, the player should keep their back in a neutral position.
- The primary spine angle should be similar to the one in the address. Leaning backwards happens sometimes when amateurs try to help the ball up in the air, and should be avoided, since it makes it harder for the golfer to aim and provide a sufficiently powerful swing [12].
- The head should be kept stable, in the late downswing, one should almost think of the head resting on a pillow and not falling backwards or forwards.

All of the above-mentioned technique features are typically observed from the side of the player, looking towards the target, in the sagittal plane.

2.2.2 X-factor and core engagement

As mentioned in Section 2.1, the X-factor is an important metric in golf performance. A technique error commonly made is to not take the time in the transition from backswing to downswing [3]. This also contributes to only using the arms in the downswing, without starting the downswing with the lower body, which decreases the X-factor. Starting the downswing with the arms is typical amongst amateurs. Similarly, beginners might only use their arms in the backswing movement, without engaging the rest of the upper body and without following the backswing movement with their shoulders. This makes them lift the club only with their arms [12].

2.2.3 Lead side lateral bending

An improper weight shift during the backswing can cause the player to lean sideways on the lead side, also called lead side lateral bending and can be seen in Figure 2.4. This puts the golfer in a sub-optimal position to start the downswing and decreases performance [4]. This technique error has also been linked with lower back pain [26] and can be observed well from the front of the player, in the frontal plane. The bending should happen towards the trailing side and progressively increase throughout the downswing, as it contributes to create the upward angle of the club head path, which leads to a higher impact speed [27]. The correct lateral bending is created as the downswing movement is started by the lower body and the upper body follows it.

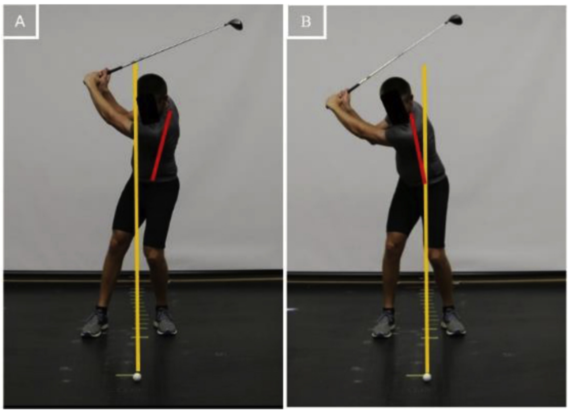


Figure 2.4: Lead side lateral bending (in A) versus proper positioning during the backswing (in B). Image: [4]

2.2.4 Swing plane

The swing plane, or the plane formed by the clubhead's movement during the swing, is also a very common cause of errors. The golfer should aim to have the same swing plane during the backswing and the downswing, and the angle between the two swing planes correlates with the ball exit direction, and a larger difference angle compromised the accuracy of the swing [3].

2.3 Injuries

Golf is not a particularly injury-prone sport, and overuse or chronic injuries are more common than acute injuries [28]. Because the swing movement places high demands on the back with high forces and torsion on the torso [4], back and shoulder injuries tend to be amongst the most common ones. Lower back pain has been reported as the most typical injury in golf [4, 29] and lower back injuries constitute up to 30% of total injuries amongst professional and recreational players [28]. Lower back pain has been linked to a lower range of motion in the hips, particularly to rotational range of motion, without a causal relationship [30]. The incidence of overuse injuries increases with the frequency of play, with the age of the golfer, and have been linked with poorer swing technique [5, 6]. This only highlights the importance of providing technique feedback to all players, especially to those with a lower skill level that might not have access to professional coaching systems.

2.4 Motion analysis

Capturing information about the pose is important in sports. Motion capture systems that rely on the use of markers, multiple cameras and specialized equipment have conventionally been used for motion analysis, and are considered to be the state of the art when it comes to capturing the kinetics of the human body. On the field, different sensors have been used to capture the kinetics of the body. For instance, Schwarz et al. [31] used Inertial Movement Units (IMU) to estimate the position of the body in golf. While sensors are more portable than motion capture systems, both are relatively equipment heavy techniques, usually require extensive knowledge about sensors and biomechanics, and are not necessarily available to recreational players. Computer vision techniques can provide simpler solutions to analyse sports performance.

2.4.1 Human Pose Estimation

Human Pose Estimation (HPE) is a set of computer vision techniques allowing to identify and classify the joints of the human body from only visual information such as a photo or a video. It consists in identifying a certain number of key points constituting the joints that are then connected to obtain a model of the human. For each joint, a set of coordinates is captured, and these joints are paired in a meaningful way, for instance, left wrist to left elbow. As opposed to motion capture systems, it does not require any markers on the human body, which makes it a much more portable and affordable option to track the kinematics of the human body in motion. Three kinds of models exist, skeleton or kinematic models, contour-based or planar models, and volume-based models, as shown in Figure 2.5. In this thesis, only the skeleton model is of interest.

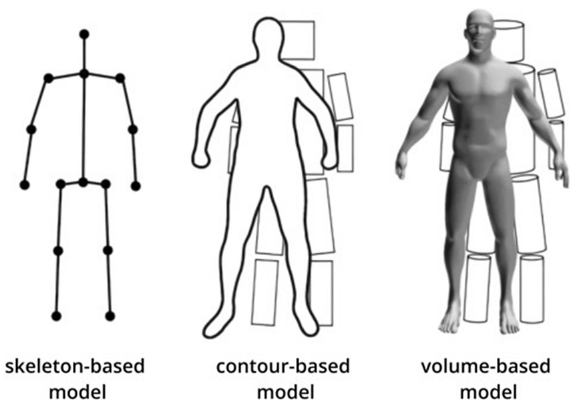


Figure 2.5: Used types of models for human pose estimation. Image: [32]

The marker-less technique using only visual image and machine learning techniques to identify the joints has gained a lot of traction in the past years. With the introduction of deep learning into the field, the models have become more generalizable to a multitude of situations and accurate compared to traditional machine learning methods. This is because of their ability to better extract patterns and features from the inputted images. In consequence, the interest and use of HPE has increased, and it is used in a large number of fields, from animation to gaming and from activity recognition to sports [8, 32, 33].

2.4.2 Human Pose Estimation in sports

In the sports field, HPE provides insights into the position and movement of the player or athlete [8]. This allows to quantify the pose of the athlete and makes it comparable both to previous performances from the same athlete or to other athletes, making it easier for the coach and the athlete to identify potential injury risks or technique improvements leading to a better performance. Especially in sports in which technique is crucial for performance, such as golf, the potential of HPE as a coaching tool is huge. The advantage of this technique is its practicality and portability, since no major equipment is required [34]. HPE can be done both in 2D and 3D, but 3D will require cameras from multiple angles or depth cameras to obtain accurate depth information. While depth cameras have gotten more common in the past years and are available on some high-end Android phones and iPhones, they are still not widespread. While 2D HPE will not provide as much information about the pose as 3D HPE would, it can still provide some valuable information about technique to the end-user.

2.4.3 Evaluating Human Pose Estimation

To evaluate the accuracy of HPE models, the joint coordinates obtained through the model are compared to the actual joint coordinates, generally manually tagged in the video. The Percentage of Correct Keypoints (PCK) metric is widely used in 2D HPE. It measures the accuracy of the localization of the human joints [33]. The joint is considered correctly detected if it falls within the threshold pixels of the ground truth joint [32]. In other words, for an estimated joint to be correct, its distance to the actual joint has to be less than a certain threshold distance. The used threshold depends on the purpose of the model and the study. Some standard thresholds are used, such as the PCKh metric, a variation of PCK, where the threshold is 50% of the head length, which makes it independent of any articulation [33, 35]. McNally et al. [25] use a temporal variant of the PCK metric in their SwingNet model. The Percentage of Correct Events (PCE) metric considers a certain event, such as the impact during a swing, correctly detected if the impact happens within a defined amount of frames. For instance, if an observer tags the event to happen in frame 5 and the model detects it in frame 6, and a margin of error of one frame has been defined, the event would have been correctly detected.

2.5 Existing solutions for swing analysis

There are multiple methods used to analyse swings. While there is some overlap, the methods used differ slightly depending on if the swing is analysed for research or on the field. The next sections will introduce the main methods used in both situations.

2.5.1 Swing analysis in research

Many studies have used motion capture to analyse the kinematics of the golf swing [3, 36, 37]. However, this method is not applicable on the field, since it requires multiple cameras, markers, and highly specialized equipment and software that can only be found in laboratories. As mentioned in Section 2.4, sensors such as IMUs have been used to analyse the swing both in a laboratory setting or on the field. Electromyographic signals have also been used to analyse the swing in laboratory conditions [38]. In research, 2D images have been used for HPE [39], but the resulting lightweight model has not been used in any commercial application to the author's knowledge. 3D HPE has also been used in research. For instance, Park et al. [34] combined images with 3D information from motion capture systems to obtain a 3D skeleton of the golfer.

2.5.2 Swing analysis on the field

The swing evaluation methods used on the field and in practice by golfers differ a little from those used in research and are described in the following sections. One of the common denominators about all the below mentioned tools is that to the author's knowledge, none of them has any public validation available. This means that while the developed tool might be evidence-based, it is not explicitly stated how the evaluation of the swing is performed and on what it is based. This project aims at filling the gap in the area.

2.5.2.1 Video

Video recording is one of the most common tool used by golfers to identify their mistakes during the swing and track their progress, but nowadays other techniques are being used as well. This includes for instance Trackman [40], which uses radar and cameras to get data about the club and the ball, including but not limited to ball speed, club speed, swing plane, spin axis and landing angle of the ball. Especially professionals use these newer tools, since they might have an easier access to them, but amongst recreational players, video is still used a lot [12]. Trackman is not the only one providing data about the swing and the ball, Toptracer range [41] also provides tools to analyse the swing, mostly from a ball tracking viewpoint. When a swing is filmed for video analysis purposes by a coach or another person, the filming angle us usually either in the frontal or sagittal plane of the player. Filming from the sagittal plane, or towards where the ball is hit, might be more common [12].

2.5.2.2 Sensors

There are portable solutions to analyse the swing that use small sensors attached to the golfer or their club. The sensor is typically an IMU placed either on the glove, like Zepp Golf 2 Swing Analyzer [42], or on the club close to the grip, such as Blast Golf Swing Trainer [43] or Garmin TruSwing [44]. All three of them provide metrics related to the speed and direction of the swing, the tempo or timing of the different phases, as well as club path and club angle, amongst others.

2.5.2.3 Mobile applications

Without the use of additional sensors, a lot of mobile applications exist to analyse the swing. A quick search on the Google Play store with the search words "golf swing analysis" yielded about 30 results, of which less than half were relevant since these results also included applications that use external sensors, score tracking applications, and entirely non-related mobile games. A lot of those applications provide insights into the swing by just allowing to capture videos from different angles, slowing and speeding them, playing them frame-by-frame, or allowing overlays to measure joint angles. Other applications offer to get feedback for the video from a coach through a subscription. The idea of using 2D images obtained from a mobile phone to analyse sports technique is not new, it has been done in many other sports as well, including running [45]. For golf, HPE has been implemented in mobile applications, but to the author's knowledge it has only been done as iOS applications and not for Android. For instance, OnForm [46] provides an iOS application with skeleton tracking, but the obtained skeleton is only used to visualise the swing, not to analyse it in any way.

Chapter 3

Methodology

An overview of this works' methodology is presented in Figure 3.1. First, a general literature research was performed about golf swing biomechanics and technique, HPE, as well as HPE specifically in golf. Then the system, starting with the pose detection and followed by the phase identification was built. After some very basic testing, it was shown that the used phase identification model was not accurate enough and an alternative way to identify the phases was developed. Following that, a mistake common in golf or a feature of the swing was chosen, a more specific literature research was done about it and an algorithm specific for that feature was developed. Then another feature was chosen, and the same steps repeated in an iterative way as many times as the time at disposal for this project allowed. Data were collected before the system was evaluated.

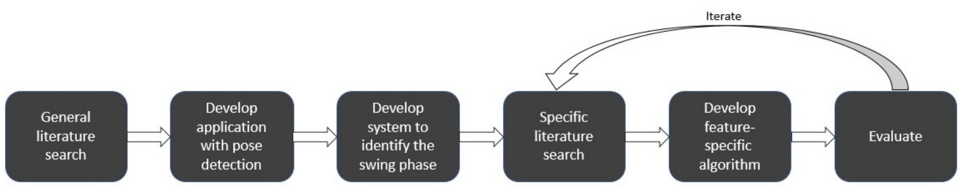


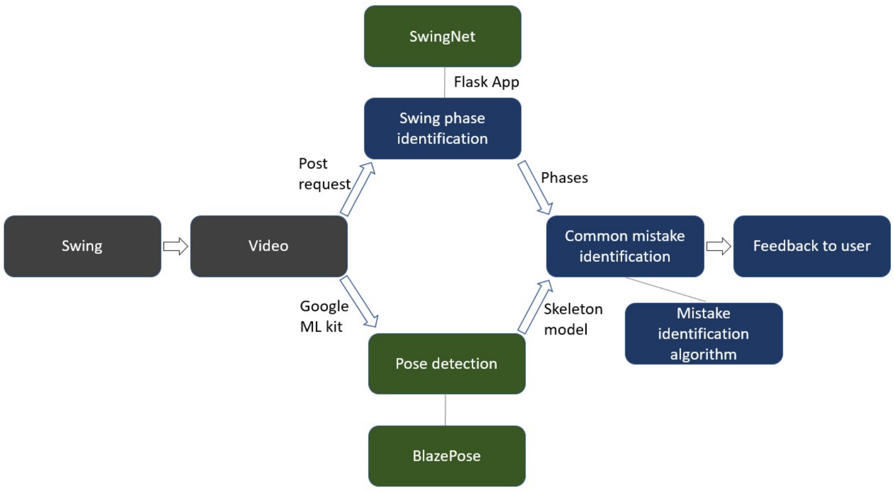
Figure 3.1: Schema of the used methodology

3.1 Literature research

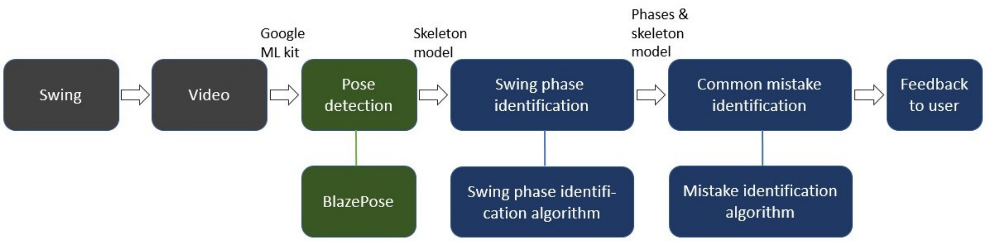
Before starting to build the system, an extensive literature research was performed. This included searches on google Scholar, PubMed, and Web of Science. The main goals of the literature search were to clarify to what extent swing phase analysis has been done, both with and without using HPE, and with what accuracy and methods it was performed. The need for filtering and extrapolation of data after obtaining the skeleton model was also clarified. Additionally, the literature search provided some knowledge in the areas of golf swing technique, injuries, and their causes, and served as a basis for discussions with golfers. This literature search also consisted of the exploratory phase of the work, where different technical mistakes and features that could be identified were considered.

3.2 Building the system

To collect all the necessary data from a video of a golf swing, a mobile application for Android, using Kotlin as a programming language, was developed. Two versions of the workflow of the application from the swing to the feedback to the user are illustrated in Figure 3.2. In Figure 3.2a an existing model is used for swing phase identification and connected through an Application Programming Interface (API) to the application. In Figure 3.2b an algorithm for swing phase identification based on the coordinates of the skeleton model is used. In both cases first the video is captured by the user, or a video is imported by the user in the app. Then, the phases of the swing were identified and pose detection is performed on the video to obtain a skeleton model of the player with the coordinates of each joint. These two types of information about the swing were combined to identify the mistakes. Two features and a technical error were selected and identified using an algorithmic approach, the specific steps for it are detailed later. Finally, the user was informed about the mistakes or lack thereof.



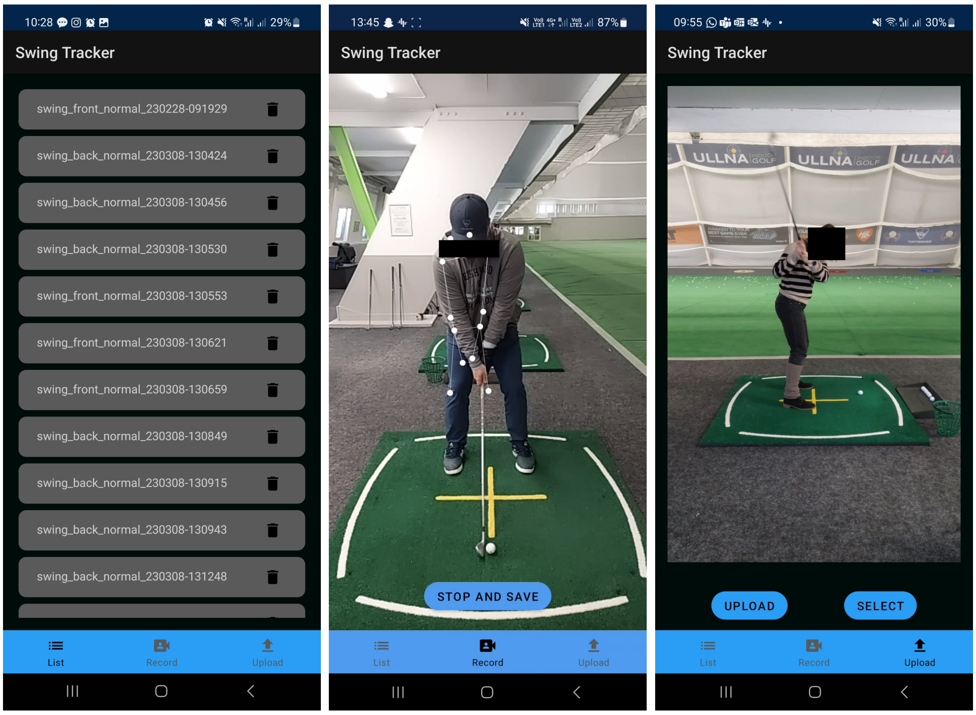
(a) Workflow of the proposed system using an existing model for swing phase identification.



(b) Workflow of the proposed system using an algorithm for identifying the phases of the swing.

Figure 3.2: Two versions of the proposed system. In dark green already existing APIs and models, in dark blue the developed elements.

The first step was to make sure that the user can both record and import videos from the local storage into the application. This was done using the CameraX Jetpack library, which aims at making camera application development easier [47]. The relatively new Photo Picker library developed by Google was used to allow the user the browse locally stored videos and import them in the application for swing analysis. Screenshots of the developed recording tool are shown in Figure 3.3. The misalignment of the skeleton overlay in Figure 3.3b is an user interface issue due to mismatching coordinates in the output of the pose estimation coordinates from the BlazePose model and the coordinates of the view, and fixing the issue was out of the scope of this project. Similarly, the key-points on the feet are not displayed on the screen but are recorded.

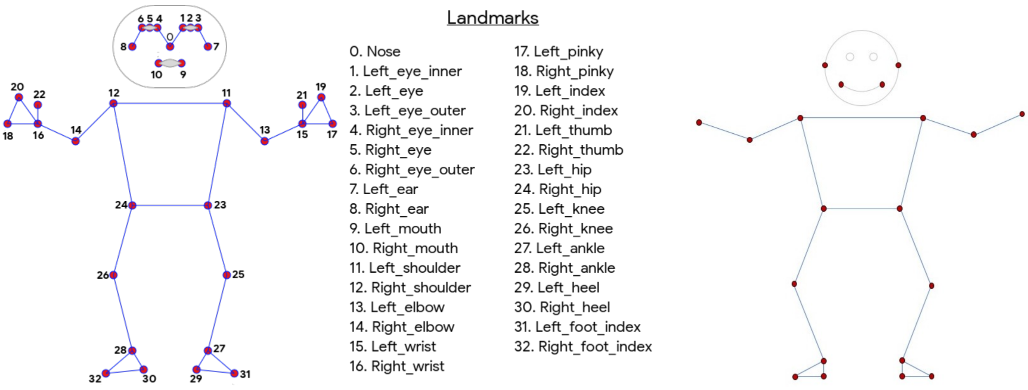


(a) Screenshot of the list view (b) Screenshot of the recording view (c) Screenshot of the upload view

Figure 3.3: Screenshots of the developed data collection application

3.2.1 Pose estimation

To estimate the pose, and already existing model with an API was used. After careful consideration detailed below, Google's ML kit's pose detection API [48] was chosen. The API is based on BlazePoze, which is a convolutional neural network that is made to run on mobile phones and is thus very lightweight [49]. The network produces 33 body key-points. BlazePose is made to estimate the pose of a human, and its limitations are that only one person can be tracked in a scene, the head needs to be visible and the person being tracked cannot be further than four meters away [50]. In this particular use case of golf swing detection, these limitations were not an issue. While BlazePose has been reported to perform worse than PoseNet, which is commonly mentioned as one of the most popular methods for human pose estimation [51, 52] and one of the models with the highest accuracy [53], it has significant advantages. Since it is much more lightweight, it can run in real-time, on a smartphone with a higher sampling rate [54]. The full BlazePose model can handle 10 frames per second and the lite model used with mobile devices 31, compared to OpenPose's 0.4 [49]. Since the golf swing is a fast movement and the feedback is wanted immediately after the swing, the possibility of having a higher sampling rate and real-time pose estimation was prioritized. Other advantages of Google's ML kit are its widespread use and great documentation, making it easy to solve any potential issues. While the lower accuracy of the BlazePose model is a disadvantage, the outputted joint coordinates had a correlation coefficient of above 0.8 with the joint coordinates given by the OpenPose model [54]. While this accuracy is not high enough for clinical settings, it should be enough for the proposed system. The ML kit's pose detection also provides a depth information for each landmark (or the z-coordinate). This however is not based on any depth camera information, but by fitting synthetic data from the so called GHUM model, another machine learning model, into the 2D data [55].



(a) Landmarks identified by the BlazePose model. Image: [49]

(b) Landmarks used in the system

Figure 3.4: Landmarks provided by the BlazePose model and the used landmarks

The BlazePose model delivers 33 landmarks, that are shown in Figure 3.4a. Out of these, shoulders, elbows, wrists, hips, knees, ankles, heels, foot indexes, ears and mouth were used. These are illustrated in Figure 3.4b. Knowing the coordinates of the thumbs, indexes, pinky fingers, nose and eye landmarks was not necessary for this application.

3.2.1.1 Filtering

The coordinates from pose detection are subject to noise, as are all measurement data. Data from pose estimation can be filtered using a Savitzky-Golay filter [23]. This is typically done to minimize the errors caused by inaccuracies in some frames of the video and smooth the joint coordinates obtained with pose detection. A Savitzky-Golay filter uses convolution to fit successive sub-sets of neighbouring data points with a low degree polynomial using the linear least squares method. A moving average filter, which when unweighted, is the simplest form of a convolutional filter, was considered. While Savitzky-Golay filters are better at preserving data features such as peak height and sudden changes [56], moving average filters are very easy to implement and effective. The coordinate position data were eventually not filtered before running the subsequently developed algorithms, because of the risks of data feature loss. In the application itself, a simple moving average filter with a window of three was implemented in Equation 3.1. This was done to have a smoother effect on the skeleton overlay. For the analysis of the swing, only raw unfiltered data were used.

$$(x_{n-2} + x_{n-1} + x_n)/3 = x_{nFiltered} \tag{3.1}$$

3.2.2 Swing phase identification

To simplify the process of analysing the swing by limiting the algorithm's use to the specific phase or moment of the swing that was investigated, phase identification was performed. Most mistakes only happen at a certain moment and in a certain phase of the swing, such as lead side lateral bending only happens in the downswing, so when creating the feature identification algorithm for that specific feature, the algorithm should only be executed on the downswing phase. In the same way, measuring knee angles is only relevant in the address and at impact, so the algorithm should only perform the calculation for those frames. In other words, the phase identification adds some context to the obtained skeleton model.

For identifying the phases of the swing, two approaches were considered. The first is an algorithmic approach that would detect some key parameters to determine which swing phase the golfer is in. This would, for instance, be the direction of the club's rotation changing between the backswing and the downswing, or the vertical position of the wrist, which changes significantly throughout the swing. The second approach is a machine learning approach using a pre-existing open source model, called SwingNet and the associated annotated videos, GolfDB [25]. Both approaches aim at identifying the frames where the so-called events happen, where an event could be for instance the impact, which separates the downswing from the follow-through. First, the SwingNet model for identifying the phases was chosen, because it was expected to provide a higher accuracy in terms of identifying the phases from different angles than an algorithmic approach would. Later this approach was changed and the reasons for it are detailed later.

3.2.2.1 SwingNet model

The open source GolfDB database provides 1400 swing videos taken from YouTube, annotated with eight investigated events in the swing sequence. These events are address, toe-up, mid-backswing, top, mid-downswing, impact, mid-follow-through and finish. The SwingNet model allows to obtain the frames at which the events happen with a reported PCE of 0.71. It is important to note that the PCE for transition between backswing and downswing, and for the impact with the ball were respectively 0.84 and 0.98, and would allow to separate the swing into the four phases mainly used in research, as mentioned in Section 2.1 with a good accuracy. The address and finish event, which marked the end of the follow-through, had a lower PCE, respectively 0.32 and 0.3, because of them being harder to precisely locate. In consequence it is not straightforward to determine the frame where the address happens compared to the impact. Nonetheless, these frames were still identified within 7 to 10 frames [25]. This accuracy should be adequate for the proposed system, considering the identification of the events is done to give some temporal sense and context to the data obtained in the skeleton model.

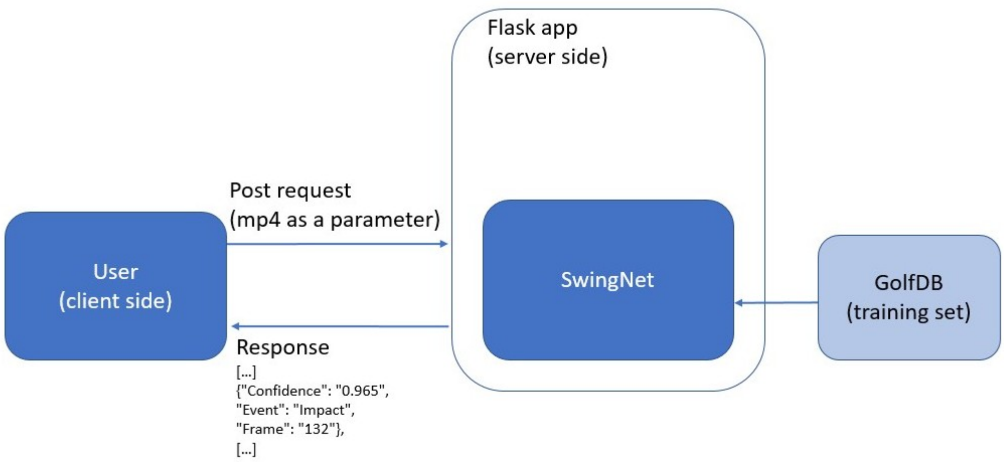


Figure 3.5: Schema of the swing phase identification server

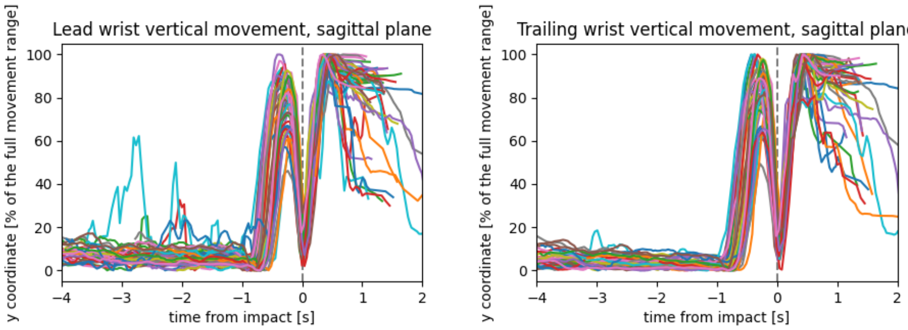
To use the Python-language SwingNet model with the developed mobile application, a restAPI was created using Flask, a micro web framework for python allowing to easily build back-ends. The developed back-end is very simple, with two endpoints: the "/" or home endpoint with information and the "/predict" endpoint, which supports POST requests with an mp4 format video, and returns a response with the event, the frame and the confidence in the format specified in Listing 3.1. The endpoint also returns the total amount of frames to account for any discrepancies between the frames extracted by the application and the server. The user can send a POST request which uploaded the mp4-file on the server. A schema of the backend is shown in Figure 3.5.

```
[
    {
        "Confidence": "0.189",
        "Event": "Address",
        "Frame": "180"
    },
    {
        "Confidence": "0.447",
        "Event": "Toe-up",
        "Frame": "402"
    },
    {
        "Confidence": "0.941",
        "Event": "Mid-backswing (arm parallel)",
        "Frame": "406"
    },
    {
        "Confidence": "0.825",
        "Event": "Top",
        "Frame": "417"
    },
    {
        "Confidence": "0.986",
        "Event": "Mid-downswing (arm parallel)",
        "Frame": "421"
    },
    {
        "Confidence": "0.988",
        "Event": "Impact",
        "Frame": "424"
    },
    {
        "Confidence": "0.971",
        "Event": "Mid-follow-through (shaft parallel)",
        "Frame": "426"
    },
    {
        "Confidence": "0.609",
        "Event": "Finish",
        "Frame": "435"
    },
    {
        "Confidence": "1",
        "Event": "total frames",
        "Frame": "473"
    }
]
```

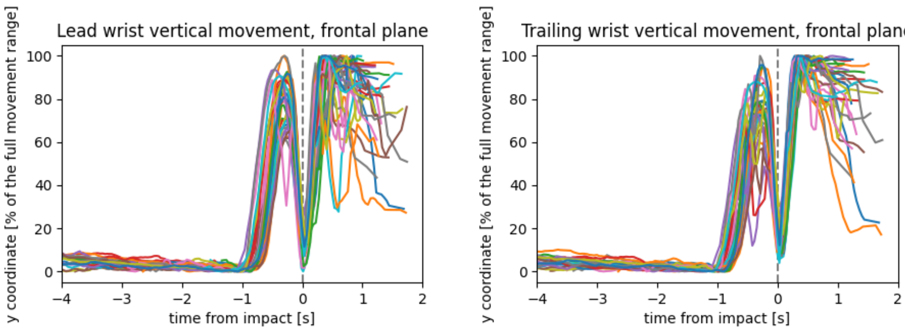

Listing 3.1: Output data from the server

3.2.2.2 Algorithmic swing event identification

After some testing, it seemed that the SwingNet model did not deliver its promised accuracy when using other videos than those provided in the GolfDB dataset. Very often it would return frames with very low confidence (below 50% likelihood of being correctly detected) and assign the events in an order that did not make sense, for instance impact before address followed by mid-downswing. This random order of the returned events was the main concern and in consequence alternative ways to identify swing events were developed in parallel with the development of the feature identification algorithms. Eventually, an alternative algorithm using the pose detection coordinates to identify the events of the swing instead of the SwingNet model was used.
Identifying the events algorithmically was done using the y-coordinate, or the vertical coordinate, of both wrists. They have a particular pattern that makes it very easy to distinguish the address, backswing and downswing from each other: there is little movement in the address, a sharp elevation in the backswing followed by a sharp descent in the downswing as can be seen in Figure 3.6. Figure 3.6a shows the vertical movement of the lead wrist (on the left) and trailing wrist (on the right) in the sagittal plane and Figure 3.6b in the frontal plane. Using only the horizontal component of the wrists has also the advantage of being filming angle independent, and the same algorithm can be used independently of whether the golfer is filmed from the sagittal or frontal plane. Both filming angles show a very similar pattern of the wrist movement, and trail and leading wrists are very similar to each other, with more or less noise. To diminish the effect of their noise while still preserving the particular pattern, the coordinates of both wrists were summed together.



(a) Vertical movement of the wrists, filmed in the sagittal direction



(b) Vertical movement of the wrists, filmed in the frontal direction

Figure 3.6: Vertical movement of the wrists over time, in the sagittal (above) and frontal plane (below), where each colored line represents one swing

As opposed to the SwingNet model which aims at identifying eight events throughout the swing, the algorithmic approach was focused on identifying three events: toe-up, top and impact. These three events were deemed sufficient for the purpose of the application and allowed to separate address, backswing, and downswing from each other. To do so, first the coordinates were transformed into a percentage of the maximal range to account for the fact that the filming might happen closer or further from the golfer, and the resulting coordinates might not have the same amplitude. Then a differentiation step was performed using the formula described in Equation 3.2, where the difference between each coordinate was taken (also called difference to previous frame in Figure 3.7), where Δwrists_y is the sum of distance moved in the vertical direction by both wrists between each frame. Because differentiation introduces more noise, Δwrist_{ver} was filtered with a moving average filter with a window of nine, which gave a good balance of preserving data features and smoothing out noise. The first half a second of recording is removed to account for eventual shaking or movements at the beginning of the recording.

$$\Delta\text{wrists}_{ver} = (\Delta\text{wrists}_y)/\Delta\text{time} \tag{3.2}$$

Toe-up

As long as the filtered $\Delta\text{wrists}_{ver}$, or $\text{MA}(\Delta\text{wrists}_{ver})$, where MA stands for moving average, was within a threshold of 3% of the total movement range, the golfer was assumed to be in the address. This threshold was chosen as it allowed for some small movements of preparation during the address, while also detecting amplitude changes quickly after they happened. This threshold also proved to be empirically better than 5 or 10%, which made the algorithm detect the events with a bigger delay and worse accuracy. Some golfers like to make some preparation movements before the actual swing, generally swing-like movements with a much smaller amplitude than the actual swing. If these movements are big enough, they can be seen such as between 0.5 and 2 seconds in Figure 3.8. The algorithm needs to be able to differentiate these preparatory movements before the address from the address. To ensure that with the above-mentioned threshold of 3% none of those preparation movements were identified as the toe-up, a condition was introduced. If no movement above 70% of amplitude was detected in the following 30 frames, the toe-up detection step was repeated five frames later. This condition was then checked again, until a movement above 70% of the full range of movement was detected within 30 frames of the detected toe-up. This frame was then chosen as the toe-up frame. After that, 10 frames, which correspond to 0.4 seconds at the frame rate delivered by the application, were subtracted as a correction. This is because that was approximately the difference in which MA(Δwrists$_{ver}$) was detected and when the toe-up actually happened.

Top

Between the top toe-up and the top, the golfer was assumed to be in the backswing. The top was identified as where MA(Δwrists$_{ver}$) changed sign. This is illustrated in Figure 3.7. Graph a) illustrates the summed vertical movement of the wrists (or the y-coordinate), and graph b) illustrates its change over time. To refine the detection of the top, a so-called top zone was defined as being the five frames before and after the identified sign change of MA(Δwrists$_{ver}$), and the frame with the highest value for the lead wrist coordinate was defined as being the top.

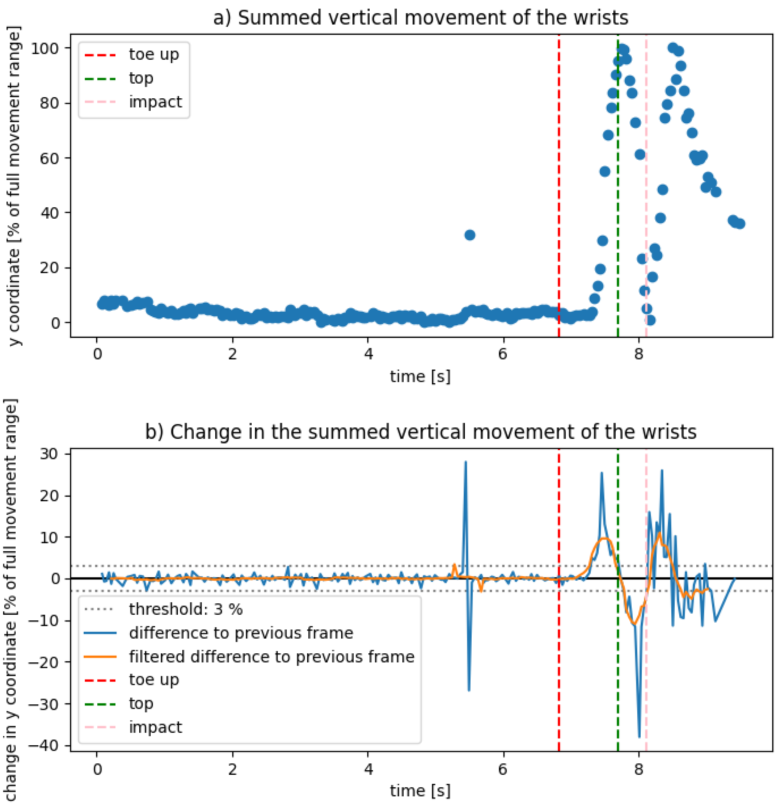


Figure 3.7: Summed vertical movement of the wrists with its change over time

Impact

The impact was detected as the moment where MA(Δwrists$_{ver}$) changes sign again. Similarly to the top zone, an impact zone was defined and the impact redefined as the frame with the lowest value of the vertical wrist coordinates. This was done by ensuring that the detected sign change was an actual sign change, and not a ripple happening around zero. To do so, Δwrists$_{ver-MA}$ at frame$_{impact}$ and frame$_{impact+2}$ were compared: if it was bigger in frame$_{impact}$, the detected event was not the impact but an unrelated ripple, and the algorithm was repeated five frames later, until frame$_{impact}$ was smaller.

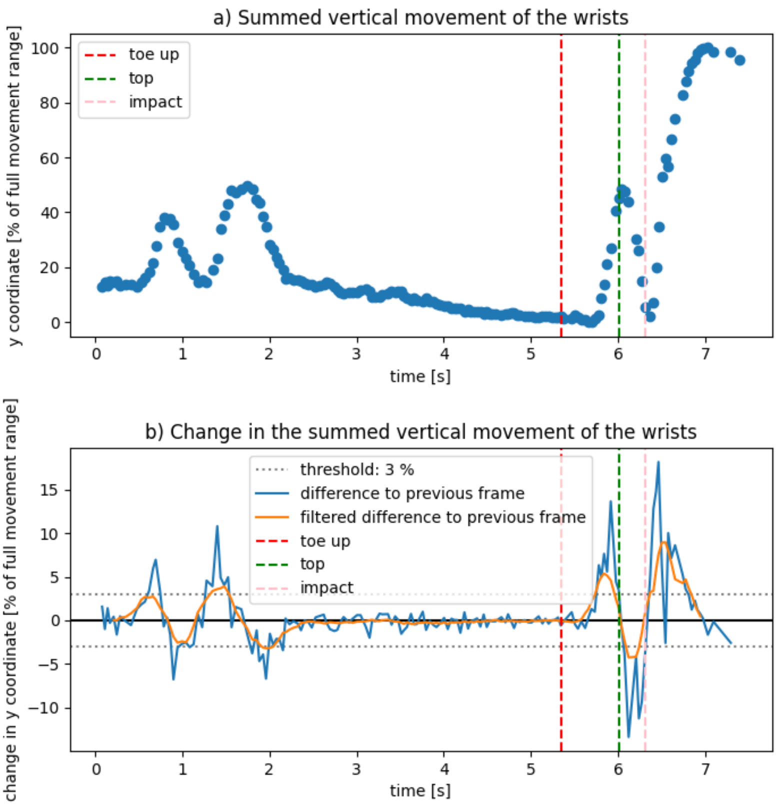


Figure 3.8: Summed vertical movement of the wrists with its change over time, with a preparatory swing-like movement between 0.5 and 2.5 seconds

Algorithm summary

The swing phase classification algorithm works in the following way:

1. Normalize the coordinates by turning them into a percentage of the full range of motion.
2. Remove of the first half a second of measurement to account for shakiness due to pressing on the record button.
3. Sum the y coordinates of both wrists together to obtain $\Delta wrists_y$.
4. Obtain the first derivative of the wrist vertical coordinates, or so called $\Delta wrists_{ver}$. In other words, how much the vertical coordinates have changed from the previous frame.
5. Smooth the first derivative by applying a moving average filter to obtain MA(Δwrists$_{ver}$).
6. As soon as MA(Δwrists$_{ver}$) goes above a set threshold, the toe-up frame can be obtained.

7. Refine toe-up to ensure it is not some preparation movement with higher amplitude by checking whether over 70% of the full movement range is reached in the following 30 frames, and then subtract 10 frames. If the condition fails, repeat five frames later, until it does not.
8. When MA(Δwrists$_{ver}$) changes sign and is now negative, the top frame can be obtained.
9. Refine the top frame by taking the frame with the highest writ y coordinate value around the previously identified top.
10. When MA(Δwrists$_{ver}$) changes sign and is now positive, the impact is obtained.
11. Refine the impact frame by taking the frame with the lowest wrist y coordinate value around the previously identified impact.
12. Ensure that the detected impact is the actual impact and not some movement that the filter has not smoothed out by ensuring that the MA(Δwrists$_{ver}$) increases in the following two frames.

In theory, the transition time could be measured from the lead wrist horizontal movement, it would just correspond to the moment where there is no movement in the lead wrist. However, the measured average frames per second of about 23 would not allow to measure transition time, which has been measured to be 0.021 seconds amongst professional golfers, as stated in Section 2.1, with sufficient accuracy, so the idea was abandoned early in the project, as it would require a higher frequency to function properly.

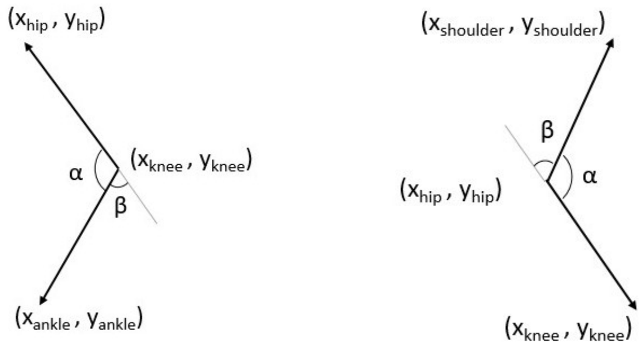
3.3 Swing analysis

An important step before selecting common mistakes to identify was to choose from which angle the golfer was filmed: either in the sagittal plane from the side of the player towards where the ball goes, or from the front of the player in the frontal plane. Not all features can be identified from both angles and in the end, filming from both angles was possible. From the known common mistakes mentioned in Section 2.2, some were selected based on their frequency and on whether 2D HPE, with its single angle limitation, could detect it. An iterative approach was adopted. Firstly, a typical mistake or feature was chosen, and additional literature research was performed about it. Then, the algorithm was drafted. Once a satisfying solution was developed, it was implemented in Kotlin in the application. Finally, the added feature was evaluated. The process was repeated as many times as time allowed. The implemented features were the following:

1. Knee flexion during the address
2. Primary spine angle during the address
3. Lateral bending angle at the top of the swing.

3.3.1 Knee flexion and primary spine angle

The first implemented swing analysis algorithm measured the knee flexion during the address. The chosen filming angle was from the side of the player, so the knee angle could be measured from the 2D image in the sagittal plane as long as there was no hip rotation. This happens in the address until the toe-up, which is the phase in which the knee angle is of relevance. Figure 2.2 shows what is the angle measured for the knee flexion, and primary spine angle.
The knee flexion was calculated the following way: $\beta = 180° - \alpha$ where α was obtained by taking the quadrant corrected arc-tangent of the hip-knee and ankle-knee vectors, as shown in Figure 3.9a. The primary spine angle was calculated in a similar way, with the coordinates as indicated in Figure 3.9b. All angles were rounded to the nearest integer. These calculations were performed for both the lead and trailing side, with the expectation of the trailing side being more accurate than the lead side, since because of the 2D image, the lead side leg is partially hidden by the trailing side leg. The angle value in the address was averaged over the ten frames before the identified toe-up frames. The ten frames correspond to approximately 400 milliseconds with the obtained frame rate, and the averaging was done to minimize the effect of any noise in a particular moment leading to the toe-up, as the address is a very static phase.



(a) Schema of the knee flexion calculation, where β is the knee flexion

(b) Schema of the primary spine angle calculation, where β is the primary spine angle

Figure 3.9: Calculation of the joint angles

3.3.2 Lateral bending angle

Since lead side lateral bending is linked with lower back pain, as stated in Section 2.2, it was an interesting technical mistake to try to detect. In this case, the golfer should be filmed with the phone facing them, from the front of the player, in the frontal plane. To identify lead side lateral bending a few assumptions were made. The phone was assumed to be held perfectly straight, perpendicular to the horizon. The location of the spine on the shoulder $Spine_{shoulder}$ and hip axis $Spine_{hip}$ was defined by taking the middle of the left and right shoulder and hip coordinates, respectively, as shown by Equation 3.4 and Equation 3.3.

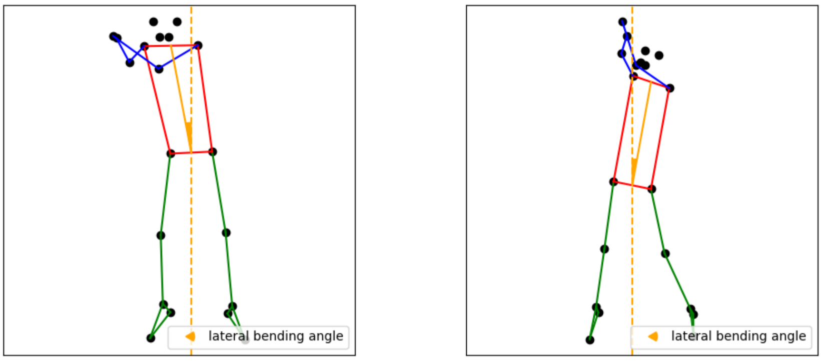
$$\text{Spine}_{shoulder} = (x, y) = \left(\frac{\text{leftShoulder}_x + \text{rightShoulder}_x}{2}, \frac{\text{leftShoulder}_y + \text{rightShoulder}_y}{2} \right) \tag{3.3}$$

$$\text{Spine}_{hip} = (x, y) = \left(\frac{\text{leftHip}_x + \text{rightHip}_x}{2}, \frac{\text{leftHip}_y + \text{rightHip}_y}{2} \right) \tag{3.4}$$

This allowed to measure the angle between the spine and the vertical line, the lateral bending angle, using trigonometry in the top frame. This angle is positive when the player is leaning towards the trailing side as in Figure 3.10a, and negative when leaning towards the lead side as in Figure 3.10b. As for the knee flexion and primary spine angle, the angles were rounded to the nearest integer.

3.4 Data collection

To evaluate the developed systems and algorithms, data collection was performed in a golf simulator located at the TopTracer main office in Stockholm and at the indoor golf installation in Ullna. The simulator allows the golfer to swing an actual ball, mimicking a real-life situation extremely closely. The advantage of performing the data collection indoors was the possibility of having a very standardized setup, with constant conditions.



(a) Skeleton model with no lead side lateral bending. (b) Skeleton with pronounced lead side lateral bending.

Figure 3.10: Lateral bending angle on the skeleton plots at the top of the swing

3.4.1 Participants

Prior to the data collection, informed consent of the participants and lack of any injuries that could prevent them from correctly swinging was ensured. The participants were women and men of all skill level, from professional golfers to complete beginners, and were all right-handed. Table 3.1 summarizes the participants. A more detailed description of the test subjects can be found in Appendix A in Table A.1. The skill level of the participants was assessed both by their years of experience in playing golf, and for those who had one, their handicap score. In golf, the lower the handicap score, the better the player. A handicap of ten and below is generally considered to be a good player. Professional golfers do not use the same system. The participants consisted of six women and nine men, all right-handed, including four complete beginners, three with a handicap over 25, three with a handicap bigger than ten but smaller than 25, and six with a handicap lower than ten. The beginners had no years of practice, and seven participants had over ten years of practice. Prior to the data collection, the participants were not instructed to wear particularly tight clothing, even though the BlazePose model is more accurate with tighter clothing [50]. This was not done because most of times, golfers might not be wearing particularly tight clothing, and the validity of the system should also be assessed taking that into account.

	N = 15	**Women = 6**	**Men = 9**
Beginner	4	2	2
Handicap > 25	3	2	1
Handicap $25 \geq x > 10$	3	0	3
Handicap ≤ 10	6	2	4
0 years of practice	4	2	2
1-10 years of practice	4	1	3
11-20 years of practice	1	1	0
21+ years of practice	6	2	4
Left-handed	0	0	0
Right-handed	15	6	9

Table 3.1: Description of the participants

3.4.2 Data collection protocol

The participants performed multiple swings and the swing was filmed from both angles: in the direction of the target, or in the sagittal plane of the player, and in front of the golfer or in the frontal plane. Participants were instructed to swing as they would in normal conditions. Six to ten swings were filmed per person, depending on their skill level. A total of 82 swings were filmed. Out of those 82 videos, four videos were left out, either because the video was not filmed perfectly perpendicular to the player, or there was not enough address in the video. 78 videos were analysed, 31 in the frontal plane and 47 in the sagittal plane. There were more videos in the sagittal plane because said plane was easier to film in the indoor golf installation in Ullna without disturbing other golfers. Both beginners and more experienced golfers were filmed in normal swings, three times in the sagittal plane, three times in the frontal plane. In addition to that, more experienced golfers were instructed to swing once with their knees as bent as possible, and once with extremely straight knees, as this would affect not only the knee flexion but also the primary spine angle, and once to fake lead side lateral bending. The participants were filmed with a Samsung Galaxy A32, and the developed application was used to obtain the joint coordinate data. The used phone was known to reach 25 frames per second when filming. When filming, the phone was held at approximately chest height, chosen so as most users would film the golfer at that height.

3.5 Evaluating the system

The accuracy of both the algorithmic event detection and the measured angles were evaluated with the methodology described below. The evaluation used the metrics PCE and PCA, as well as common statistical analysis methods such as mean and standard deviation. The validation was performed using a script written in Python.

3.5.1 Algorithmic swing phase identification

To assess the accuracy of the developed swing phase algorithm, the event frames given by the algorithm and the event frames obtained by annotating the video were compared. To ensure reliability and reproducibility, the frame number for the toe-up, top, and impact frame were written down three independent times and then the average value of the three was taken. The frame number of the video was obtained using the software Adobe Premiere Pro and the events were considered in the following way:

- Toe-up: the toe-up frame was considered to be the last frame before the club starts moving.
- Top: the top frame was the frame where the club was reaching its furthest and highest position.
- Impact: because of the relatively low number of frames per second filmed by the phone camera, the video rarely contained a frame of the club hitting the ball. In the cases where that frame didn't exist, the last frame before impact was considered to be the impact frame.

As the frames per second are relatively low, the investigated event might happen in between two frames and it is up to the person annotating the videos to decide in which of the frames the event happens, an error margin of one frame was introduced for all events. This means that if the algorithm detects the toe-up frame to happen in frame 23, and the observer notes down that the toe-up happens in frame 24, the toe-up is considered correctly detected, as the person annotating the video might have had to choose between frame 23 and 24, as the impact might have happened between the two frames. As stated in Section 3.2.2.1, the SwingNet model uses PCE to evaluate the accuracy of the model, and it was used to assess the accuracy of the developed algorithm as well. Three other metrics derived from it were used: PCE1, PCE2 and PCE5. The number states the tolerance, for instance, PCE2 considers frames detected within two frames of the actual frame as correctly detected.

3.5.2 Angle detection

Ideally, the best way to evaluate the accuracy of the system for knee flexion and primary spine angle would be to compare data from state-of-the-art motion capture systems. This, however, was not applicable due to the lack of equipment. The use of a goniometer was also considered, but because of the inherent dynamism of the swing movement, it was not applicable. As seen with the PCK metric, assessing the accuracy of HPE models is very often done by comparing the ground truth joint and the detected joint, where the ground truth joint is annotated in each image [8]. The same method can be used to assess the accuracy of angle measurements, which are based on the joint coordinates. The evaluation of the knee and primary spine angles was done by measuring the angles from the address frame of the video using the open-source software Kinovea, a video analysis tool designed for sport analysis. The lateral bending angle was measured in the same software, but by taking the angle of the spine line described in Section 3.3.2 to the vertical during the top frame. While the developed algorithm outputs the hip flexion and primary spine angle for both trailing and lead sides, only the trailing side angles were validated using this method, because of the inaccuracy of measuring angles of the joints hidden by other joints. Validation of the angles on the lead side will be left as a further work that would include results from motion capture systems or cameras in multiple angles. To assess the accuracy of the angle, a variation of the PCE metric, called Percentage of Correct Angles (PCA) was used, giving out the amount of correctly detected angles out of all angles. An error margin of 1% was set. This means that for the angle to be properly detected, it needs to be within 1% of the actual angle. In degrees, it means that a discrepancy of up to 3.6 degrees is considered correctly detected. A second metric, PCA2 was used, and this metric assumed an error margin of 2%.

3.5.2.1 Knee flexion and primary spine angle

The main challenge was to annotate the angles in a standardized way across all videos. To ensure reproducibility and reliability, the points to measure the angles were placed in the centre of the joint and an additional description of the joint location is provided in Table 3.2. Each angle was measured three independent times, and the average of the three angles was taken.

Joint	**Location**
Ankle	On the lateral malleolus
Knee	In between the front and the back of the knee
Hip	In between the front and the back, on the femur's protrusion
Shoulder	On the acromium

Table 3.2: Description of the joint locations used for annotating the videos

3.5.2.2 Lateral bending angle

Similarly to the evaluation of the previously mentioned angles, knee flexion and primary spine angle, the lateral bending angle was evaluated comparing manually measured angles at the top of the swing and the value returned by the algorithm. The angle was obtained by taking the angle to the vertical of the spine line, from the middle of the hips to the middle of the shoulders.

3.5.3 Reliability and validity of the data and methods

Measures were taken to ensure the validity and reliability of the data. The points at which particular attention was paid and the known pitfalls of the used methodology are listed below.

3.5.3.1 Reliability

To ensure test-retest reliability when collecting the data, multiple swings were filmed per participant, on multiple days, in two different locations. To assess the reliability of the data collection, performing the data collection another time and comparing the outcome of it to the first one was considered, but not performed due to time concerns. When measuring the angles from the video, with the methodology mentioned in Section 3.5, the measurement was repeated multiple times to improve the test-retest reliability. Inter-rater reliability was not ensured, and in further validation steps, different people should measure the angles from the video, to minimize one person's potential bias.

3.5.3.2 Validity

To ensure that the collected data was applicable to any golfer, particular attention was paid during the data collection so that the participant group was composed of golfers of all skill levels. Both data from women and men was collected, with a ratio of 2:3. These steps aimed at ensuring that the collected data was valid enough to be generalized to most golfers. Further validation steps should include a bigger number of participants. Elderly golfers and golfers with injuries or disabilities were underrepresented in the test subject group.

Chapter 4

Results

This chapters presents the main results of the validation of the developed algorithms. Complete results for the phase identification can be found in Appendix B. A total of 78 videos were analysed, of which 47 were filmed in the sagittal plane and 31 in the frontal plane. Phase identification was evaluated for both filming angles, and angle identification was evaluated for their relevant filming angle. The frames per second delivered by the application were not constant and were on average 23.15 frames per second with a standard deviation of 5.49 over the whole dataset.

4.1 Output of the developed algorithms

The developed algorithms allow the user to obtain the frame or time from the start of the video to the events toe-up, top and impact. If the video is filmed in the sagittal plane, the knee flexion and primary spine angle is obtained, and if the video is taken in the frontal plane the lateral bending angle in the top is obtained.

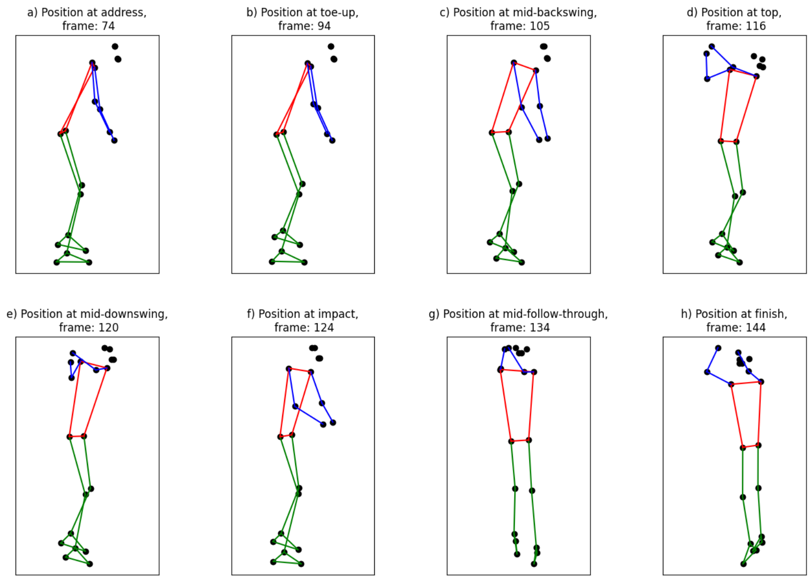


Figure 4.1: Skeleton plot of the golfer at moments of interest throughout the swing, in the sagittal plane

The output of the event identification allows to plot the position of the player throughout the swing, as for instance in Figure 4.1. There, the golfer is filmed in the sagittal plane. The event detection algorithm allows to obtain the frame for the toe-up, top, and impact. In Figure 4.1, the address frame is chosen as the frame 20 frames before the toe-up, the mid-backswing is the frame in the middle of the toe-up and top, the mid-downswing the frame in the middle of the top and impact, the mid-follow-through the frame in the middle of the impact and the last frame, and the finish frame was the last frame of the filmed swing. More sequences like this, including in the frontal plane can be found in Appendix C.

4.2 Used metrics

The results are presented using the PCE and PCA metrics presented in Section 3.5.1 and Section 3.5.2. The accuracy of the algorithm was also evaluated using mean errors with their standard derivations. It is important to note that both mean errors and absolute mean errors were used, where the mean absolute error is the mean of the absolute error values. Mean errors were mostly used to assess if the algorithm had some drift and to quantify its precision. For instance, it was meant to see whether the knee angle was systematically detected lower than it actually was. Absolute mean errors, which did not consider whether the error was positive or negative, just that there was an error, were used to assess the accuracy of the algorithm.

4.3 Event identification

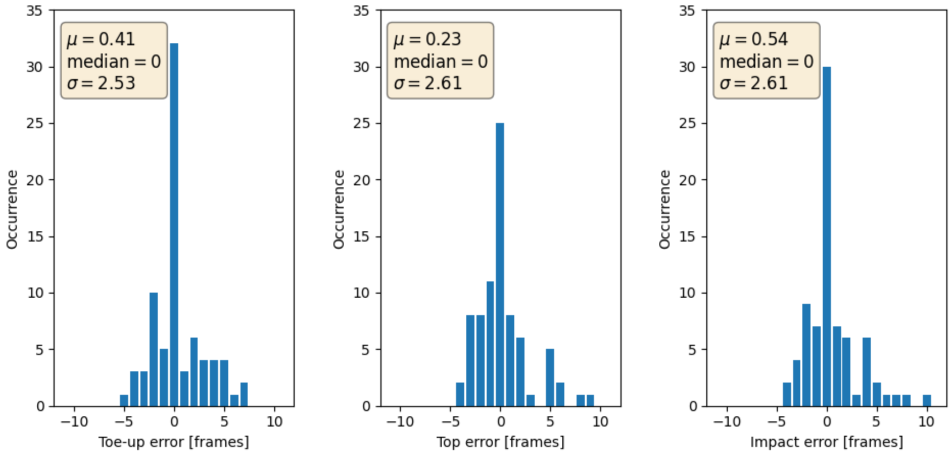
The accuracy of the event identification was assessed for the event detection algorithm developed in the scope of the project, as well as for the SwingNet model, as a comparison.

4.3.1 Algorithmic detection of the phases

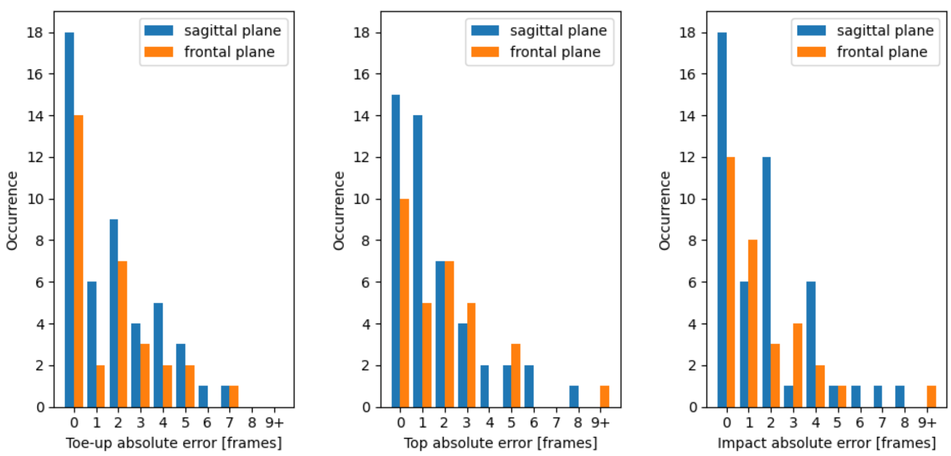
Filming angle	**PCE**	**PCE1**	**Mean absolute error in frames**
Sagittal	0.39	0.55	1.7 ± 1.9
Frontal	0.36	0.55	1.7 ± 2.0
Sagittal & frontal	0.37	0.55	1.7 ± 2.0

Table 4.1: Overview of the results of the event detection algorithm

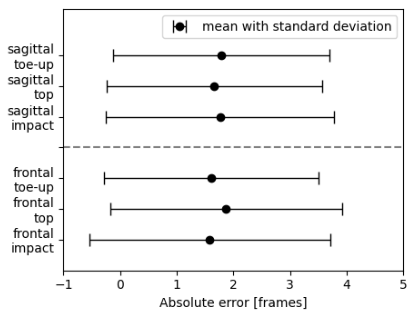
The proposed algorithm succeeded at correctly detecting the event frames in 37% of the cases. The event was correctly detected within one frame in over 50% of the cases, with a PCE1 of 0.55. Within five frames of the actual event, 95% of the events were detected. With the average frame rate of 23.15 frames per second reached by the hardware, this means that 95% of the events are detected within 217 milliseconds of the actual event. The mean absolute error for detecting all events was 1.7 ± 2.0 frames for the videos filmed in the frontal and sagittal plane. As can be seen in Table 4.1, the accuracy of the algorithm did not differ significantly depending on the filming angle and was very similar in both planes, respectively a PCE of 0.36 and 0.39, and a PCE1 (with a one frame tolerance) of 0.55 in both cases. Table 4.2 shows additionally that while there are some minor differences for each event, there are no major differences in how well the different events are identified, with the toe-up and the impact being detected slightly better than the top, at least when only factoring PCE.



(a) Error distribution for each event, with their mean (μ), median, and standard deviation (σ)



(b) Absolute error distribution for each event by their filming angle



(c) Absolute error with the standard deviation for each event by their filming angle

Figure 4.2: Error and absolute error distribution for the developed event detection algorithm

Plane	Event	PCE	PCE1
Sagittal	Toe-up	0.38	0.51
	Top	0.32	0.62
	Impact	0.38	0.51
Frontal	Toe-up	0.45	0.52
	Top	0.32	0.48
	Impact	0.39	0.65

Table 4.2: Overview of the results of the event detection algoritm by plane and event

This can also be seen in Figure 4.2. Figure 4.2a shows that the distribution, mean, median and standard deviation of the errors in toe-up, top and impact are all similar to each other. Additionally, there is no significant systematic error, as the mean and median error values are in between 0 and 0.6. Figure 4.2b shows how the absolute errors are distributed. Most of the the videos have errors that are smaller or equal to two frames, as the PCE1 for all events is of 0.74. The figure highlights as well how most of the events are identified within five frames of the actual event. Figure 4.2c illustrates the mean absolute error with its standard deviation for each event in both filming angles and shows that there is no significant difference between the events and filming angles.

4.3.2 Results of the SwingNet model for comparison

Event	PCE	PCE1	Mean absolute error in frames
Toe-up	0.48	0.52	1.7 ± 2.1
Top	0.32	0.62	2.1 ± 2.2
Impact	0.33	0.52	1.8 ± 1.9
All	0.38	0.51	1.8 ± 2.0

Table 4.3: Results by event for the event frames predicted by the SwingNet model

To compare the results of the algorithmic event detection to the existing model SwingNet, 21 videos were evaluated. The accuracy is very similar to the accuracy of the developed algorithm, with a PCE of 0.38. The accuracy is better for the detection of the toe-up frame, 46% compared to the 32% and 33% of the top and impact. Similarly to the developed algorithm, 95% of the events were detected within five frames. The mean absolute errors shown in Table 4.3 shows that there is no significant difference in the mean absolute error compared to the developed event detection algorithm, as the errors are all in a similar range.

4.4 Angle measurements

The knee flexion, primary spine angle and lateral bending angle algorithm were validated. The angle measurements in the sagittal plane (knee flexion and primary spine angle) were validated only for the trailing side, as the chosen validation method prevented from doing so for the leading side.

4.4.1 Knee flexion and primary spine angle

Table 4.4 displays the main results of the angle identification during the address in the sagittal plane. The measurement of the knee flexion exhibited an accuracy of 66% within the error margin of 1% or 3.6 degrees. This accuracy reached 85% if the error margin was doubled. The primary spine angle was detected with a much lower accuracy of 4%.

Angle	**PCA**	**PCA2**	**Mean absolute error**
Trailing knee flexion	0.66	0.85	$3.7° \pm 3.8°$
Primary spine angle	0.04	0.32	$10.5° \pm 5.0°$

Table 4.4: Overview of the results of angle measurements

The detected angles tended to be bigger than the true angles, as can be seen in Figure 4.3. This was particularly true for the primary spine angle, as the mean error was 10.32 degrees, which was less accurate than the knee flexion, with a mean error of 2.98 degrees. The precision of the primary spine angle was slightly worse than the knee flexion, as the standard deviation of the mean error is higher for the primary spine angle.

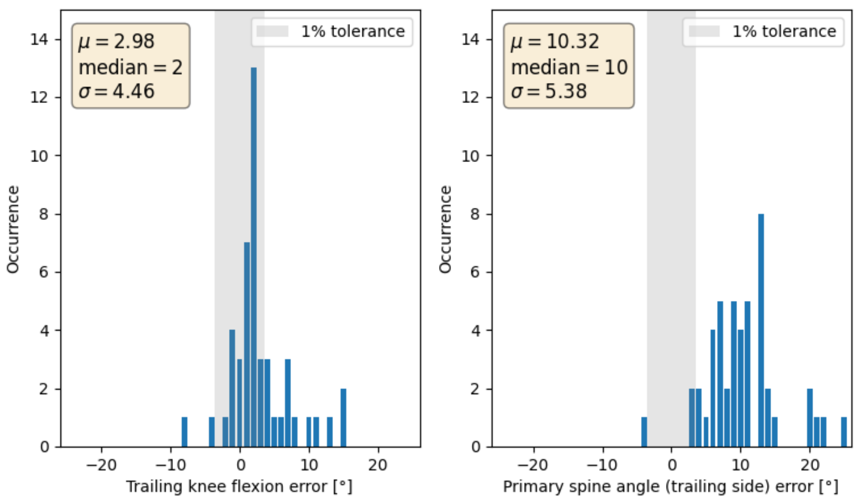


Figure 4.3: Error distribution for the knee flexion and the primary spine angle on the trailing side

In Figure 4.3, a small outlier group with knee flexion error values between 10 and 15 degrees, and primary spine angle error values above 20 degrees was identified, which belonged to the same five videos.

4.4.2 Lateral bending angle

In the 31 videos taken in the frontal plane, lateral bending angle was identified with an 87% accuracy. With a higher tolerance margin of 2% or 7.2 degrees, the accuracy reached 100%. Table 4.5. Figure 4.4 shows how the errors were distributed. The detected angle tended to be

Angle	PCA	PCA2	Mean absolute error
Lateral bending angle	0.87	1.00	$1.6° \pm 1.4°$

Table 4.5: Lateral bending angle results

slightly bigger than the actual angle, as the average error was -1.29 degrees.

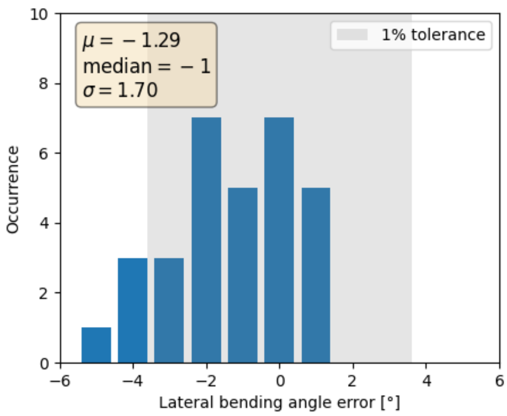
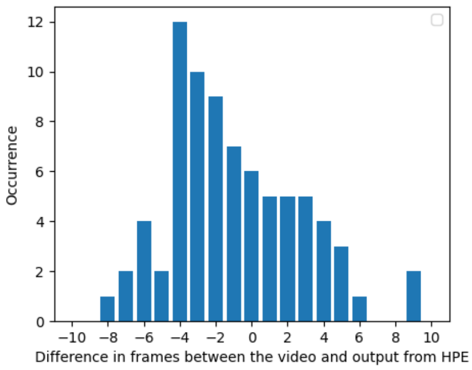


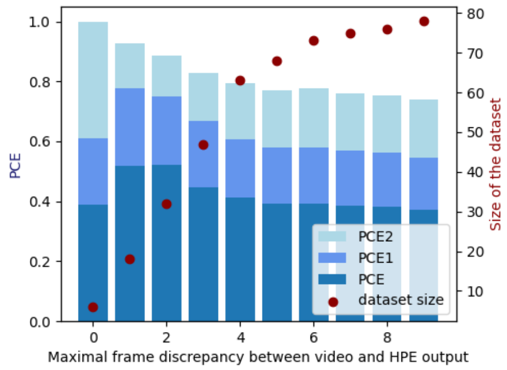
Figure 4.4: Error distribution for the lateral bending angle

4.5 Number of frames in the original and processed video

To assess if the processing of the data affected the number of frames of the video, the total amount of frames after the pose detection was performed were compared to the total amount of frames of the original video. As shown in Figure 4.5a, the frames on which pose detection was performed did not perfectly correspond to the number of frames in the original video. A negative value indicates that the original video had less frames than the processed one, and a positive indicates that the original video had more frames than the processed one.
The discrepancy influenced the accuracy of the event detection, with a lower discrepancy in the total number of frames being generally more accurate. Figure 4.5b shows the effect of the discrepancy on the PCE metric and the size of the dataset. The size of the dataset was at its biggest at 78, and at its smallest at 6, when only videos where the total amount of frames was identical in the original video and the processed one. The PCE metric tended to be higher the smaller the discrepancy, apart from a discrepancy of zero. A discrepancy of maximum one seemed to be the more accurate, with a PCE of 0.52.



(a) Difference in the total number of frames between the original video and the frames on which pose detection was performed



(b) Effect of the difference in the amount of frames to the PCE metric, with the size of the dataset

Figure 4.5: Frame discrepancy and its effect on the used metrics

Chapter 5

Discussion

This chapter aims at discussing the obtained results. It introduces the limitations and the potential sources of errors of the used methodology. Some of the improvements that could be done and future works are mentioned.

5.1 Filming angle independence of the event detection algorithm

As described in Section 4.3, the accuracy of the phase detection algorithm was very similar when filming in the sagittal and frontal plane, both with a mean absolute error of 1.7 ± 2.0 frames. This highlights the potential of using the vertical movement of the wrists during a golf swing as a filming angle independent way of detecting events and phases during the golf swing.

5.2 Algorithmic approach versus machine learning for event detection

Interestingly, the accuracy of the event detection is very similar for both the developed algorithm and the SwingNet model. The SwingNet model performed far worse than expected, by delivering at best a 0.48 PCE compared to the 0.84 promised by the paper [25]. While these results of the validation of the developed algorithm and the SwingNet model are not fully comparable to each other as only part of the data collected for the evaluation of the developed algorithm was used to evaluate the performance of the SwingNet model, the findings give a good indication of what kind of accuracy to expect. Reasons for this difference between the obtained and expected results of the SwingNet model could be due to different frame rates. The videos used to train the model were mostly taken outdoors compared to the indoors videos filmed in the scope of this project, and they might have different luminosity conditions. Part of the training videos were slow motion videos, but it is unknown what kind of effect this would have on the accuracy. The SwingNet model was also trained on relatively few videos, and the model might have been over-fitted. An optimisation of the parameters of the model could be performed. While eventually, despite of the apparent temporal disorder in the returned events, the model did not perform significantly differently than the developed algorithm. This could indicate that one of the ways to have a batter accuracy with the SwingNet model would be to modify the model to ensure that the events get identified in the correct temporal order.

5.3 Source of errors

Some of the most important potential sources of error are presented, and ways for preventing them from happening are listed below.

5.3.1 Eventual discrepancies in the BlazePose model

As presented in Section 4.5, the number of frames in the original video did not generally match the amount of frames processed by the BlazePose model. While it was expected that some frames would be lost due to the processing of the pose potentially taking some time, it was not expected of the BlazePose model to add frames in the processing.

5.3.1.1 Video versus skeleton model

What typically stood out in the videos with a high error in the event detection was that they also had a high discrepancy in the number of frames between the original and processed video. This can be illustrated well by plotting the skeleton model for the nine frames surrounding the detected event (four before the event, the event itself, four after the event), in Figure 5.1. The position of the golfer in the original video, Figure 5.2, does not match the skeleton plot at that same frame.

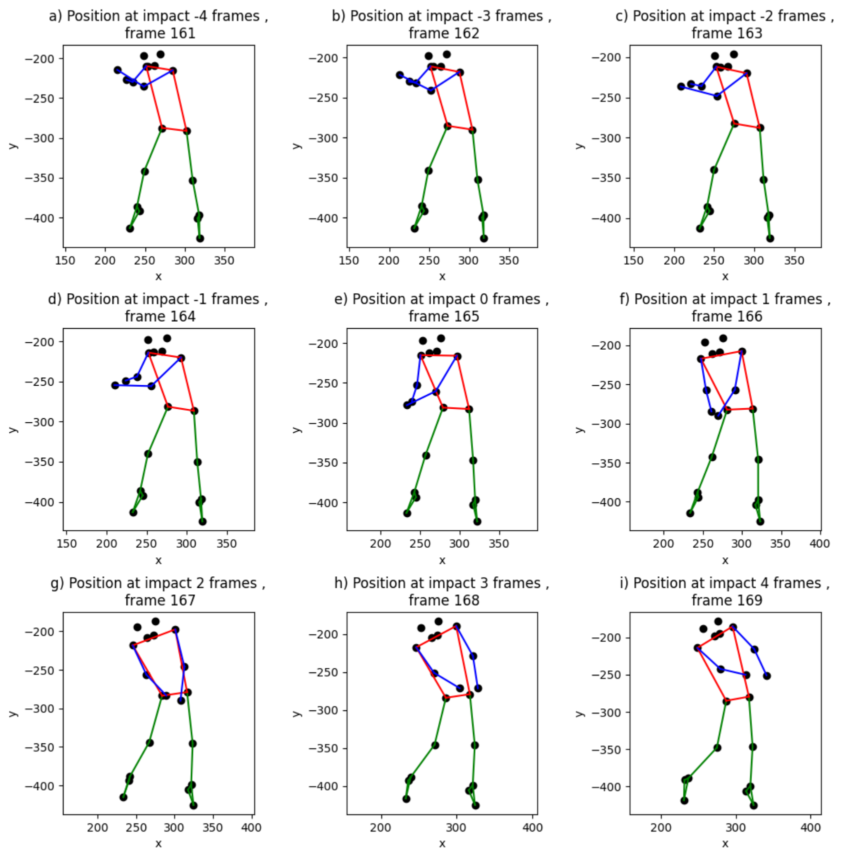
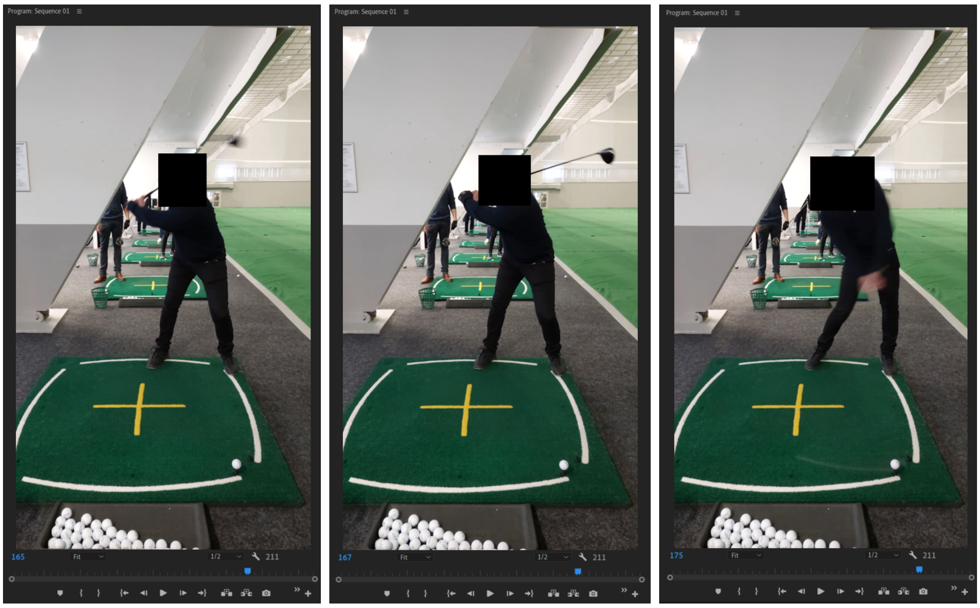


Figure 5.1: Frames around the detected impact of a swing, showing potential discrepancies in the BlazePose model

In Figure 5.1, the impact is detected by the algorithm in frame 165, whereas visually it would look like the impact happened two frames later, in frame 167. On the original video however, as shown in Figure 5.2, the impact is detected in frame 175. Frame 165, which is detected by the algorithm as the impact, actually seems to be before the top of the swing, as illustrated by Figure 5.2a.



(a) Frame 165

(b) Frame 167

(c) Frame 175 with the ground truth impact

Figure 5.2: Frames of the video

This discrepancy was not visible for every video and was this marked only in very few of the videos. These discrepancies however exist and would suggest that validation by comparing frames might not be the most optimal way. Further work should be conducted to investigate the causes of this discrepancy and its effect to the accuracy of the algorithm. A timestamp instead of a frame number might be better at matching the pose and the video frame.

5.3.1.2 Effect on the accuracy of the presented algorithms

While the accuracy of the algorithm seemed to be better in the videos with a smaller discrepancy, as mentioned in Section 4.5, there are multiple reasons why these better accuracies might not be representative of the reality. First, the dataset size for no discrepancy in frames number or one frame is very small, respectively 6 and 18 videos. Secondly, there might still be discrepancies in, for instance, the first half of the video that get compensated in the second half. For example, the processing could skip a frame before the top, but add a frame after the top. This cannot be assessed with the available data and requires further investigation. This discrepancy had a much smaller effect on the feature detection algorithms. This can be explained because the knee flexion and primary spine angle were averaged over ten frames. For the lateral bending angle, the obtained angles are very small, ranging from -13 to 13 degrees. The lateral bending angle will also generally be very similar in the five frames preceding and following the top of the swing, so a perfect detection of the top frame is not necessary to detect whether there is lead side lateral bending or not.

5.3.1.3 Real-time processing

This discrepancy might also be purely a real-time processing issue, as The BlazePose model API does not queue up frames if it doesn't have time to perform pose detection on the sent frames.

However, this does not explain where the additional frames compared to the original video come from. It remains to be seen if the issue would persist if the pose detection coordinates data had been obtained in post-processing, by uploading the video instead of calculating the pose while the filming happened, as it was done in the data collection.

5.3.2 Inclination of the phone

For the various phase detection and swing analysis algorithms to work properly, some assumptions are made. The algorithm assumes that the phone is held perfectly perpendicularly to the floor, facing the filmed golfer either from the frontal or sagittal plane, and that the phone is held perfectly stable. The coordinates obtained for each joint are relative to the position in which they are projected on the phone screen. If the phone is shaken, the algorithm will assume that the person is moving, and if the phone is held with a five degree lateral inclination, the vertical will be inclined by five degrees from the actual true vertical. Similarly, if the phone is held too high or too low compared to the golfer, the measured knee flexion and primary spine angle might be different due to parallax error and the apparent displacement of an object, or in this case an angle, viewed from different positions. A validation using 3D motion capture systems or goniometers could be done in the future to assess the effect that the position of the phone camera relative to the golfer has on the angle measurements. As angle measurement during a golf swing in research has mostly been done using 3D motion capture systems, and, to the author's knowledge, there have been no attempts at quantifying the angles investigated in this thesis with alternative methods, it is not possible to compare the developed algorithms to existing different strategies. This only highlights the need for further research in the matter before a comprehensive validation can be performed.

5.3.3 Joint centres

As seen in Section 4.4, the angle measurements are more accurate for knee flexion and lead lateral side bending than for the primary spine angle, while the primary spine angle only has a slightly lower precision than the knee flexion detection. This is believed to be caused by the pose detection model not assessing the shoulder at the same place as the manual validation step described in Section 3.5.2.1 did. Bazarevsky et al. [49] did not mention the exact localisation of the ground truth joints in BlazePose model. Using the same location would likely increase the accuracy of the developed primary spine angle detection algorithm by removing the drift, as the standard deviation of the error, 5.38 degrees, is not much bigger than the standard deviation of the knee flexion algorithm, 4.46 degrees. While the effect seems to be minimal on the knee flexion and primary spine angle, using the same joint localisation could improve the validation and reflect more the reality. Additionally, when tagging the videos and measuring the joint angles from there, even though a rigorous protocol consisting of multiple measurements and clearly defined joint centres was applied, errors might have occurred in the validation step.

5.3.4 Clothing

As mentioned in Section 4.4.1, there was a small group of outliers with a big discrepancy between the ground truth angle and detected angle, both in the knee and hip joint. These five swings were all performed by people wearing looser clothing. This made assessing the joint centre harder for the manual validation, and probably caused some errors in the pose detection as well, since looser clothing is a possible error factor when using the BlazePose model [50]. The validation step could be improved by having more standardized conditions: for instance by having test participants wear clothing with markers on the joint centre, to make it easier to identify joint angles from the video. Similarly, a more reliable validation could be executed by using motion capture systems in a laboratory setting. In the case where the validation is performed with tighter clothing, how looser clothing affects the accuracy and precision should also be assessed.

5.4 Potential improvements and further work

As the goal of this project was to investigate the possibility of using 2D HPE on a mobile phone to analyse a golf swing, multiple potential improvements and further work can be mentioned.

5.4.1 Improvements to the existing system

The existing system could be improved in multiple different ways. The more interesting ones are listed below.

- **More features and mistakes**: The system developed in the scope of this project did not aim at being exhaustive in what can be detected with 2D HPE. Currently, it only allows to measure three different angles on the player. As such, the system would be rather useless for a recreational or professional player, or a coach. The logical next step would be to identify more features or mistakes that can be detected. Some are mentioned below, but more are possible.
- **Primary spine angle at address**: Measuring the primary spine angle at impact was considered at the beginning of the project but abandoned due to the low accuracy of the primary spine angle measurement during the address. If the drift in the primary spine angle measurements were to be corrected by taking the same joint coordinates as in the pose detection model, as stated in Section 5.3.3, the proposed system could be adapted very easily to detect the primary spine angle at impact as well.
- **Refining phase identification with other joint coordinates**: Only the wrist coordinates were used to identify the swing events in the developed algorithm. In an extension work, the developed algorithm could be refined using more than just the wrist coordinates for it. One could use the moment where the wrist vertical coordinates pass above the hip, elbow, or shoulder to identify further events in the swing. Similarly, the used thresholds could be improved for increasing the accuracy.
- **Thresholds for all frame rates**: The currently used thresholds assume that the captured videos have about 23 frames per second, which is the frame rate reached by the developed system and limited by the used hardware, a Samsung Galaxy A52, which is a mid-range phone from 2021. It is currently unknown whether the algorithm works for different frame rates. It should both be tested and adapted if necessary. For instance, with a higher frame rate, the moving average window used to filter the difference in vertical movement between the frames might need to be adapted. With a newer high-end phone with better computational power, the frame rate would probably be higher. This would allow to fine-tune the detection of the events and, depending on the reached frame rate, could make the detection of the transition time possible.
- **Phone orientation and movements**: Another way of refining the algorithm would be consider the orientation of the phone in space using the phone's built-in sensors. This would allow to correctly detect the lateral bending angle even if the phone was held sideways. In a similar way, the potential noise in the coordinates coming from shaking the phone could be minimized.
- **End-user perspective and needs**: This work only aimed at investigating the feasibility of using 2D HPE on a mobile phone to give feedback to a golfer. If this were to be implemented with a commercial application in mind, end-user tests would have to be performed to investigate the needs of professional and recreational golfers, as well as coaches. More features that could be investigated with this system should also be identified.
- **From quantitative to qualitative data**: With the existing system, only quantitative data, the angle measurements, are obtained. In a future work, acceptable or correct thresholds for the angles should be defined. For instance, for the lateral side bending, a threshold above which the bending is considered a mistake should be defined, so that the user could

get feedback as to whether they are leaning too much on the lead side. This would require extensive work and research into defining what is a technically correct swing.

- **Validation with motion capture**: A logical next step into ensuring the correctness of the system would be to compare the obtained joint angles with those returned from a motion capture system. This would allow to assess how big of an effect the height at which the phone is held has on the measured angles. This kind of comparison to 3D data would allow to estimate the effect of a parallax error on the measures obtained by the developed system. Additionally, this would allow to validate the detected angles on the lead side, which are partly occluded by the trailing side.
- **Clubhead detection**: In Section 2.2.4, the swing plane was mentioned as one typical source of errors golfers make. This could be detected as well in a pose detection model to analyse the golfer's swing plane. It would require training a new model where the club's head would be detected in a similar way as any joint is. This would require extensive work into obtaining and annotating a dataset. Alternatively, a lightweight IMU sensor could be placed on the club-head, and would require sensor fusion to combine the inertial data from the IMU and the kinetic data from HPE. Detecting the clubhead would allow to obtain various data about the swing that cannot be provided by the joints only.

5.4.2 Further work with the BlazePose model

The BlazePose model provides a depth coordinate, or z-coordinate as well, in addition to the x and y-coordinate used in the proposed system. As stated in Section 3.2.1, the third coordinate is based on synthetic data from the GHUM model, another machine learning model. The presence of the 3D information would allow to get information from the swing such as the X-factor, the swing plane, and timing of the shoulder axis and hip axis rotations. However, since the z-coordinate data are based on synthetic data, an extensive validation with 3D motion capture systems would have to be performed.

5.4.3 Further work including machine learning

Besides improving the developed algorithm, further work could investigate the use of the obtained joint coordinates in a machine learning model. As stated in Section 3.2.2.2 and visualized in Figure 3.6, the horizontal movement of the wrists has a characteristic pattern. This is also true for many other joint coordinates, and while those may be more difficult to interpret with a human eye, they seem to be suitable for training a model. A deep learning network might be better at considering the coordinates from all joints to identify the events than an algorithm would, without being too complex to develop. Additionally, it would be interesting to compare the performance of the system if OpenPose had been used instead of BlazePose. As BlazePose is slightly more accurate than OpenPose, the algorithm's accuracy could have been improved, but whether that would come with drawbacks in terms of having to wait for the processing of the video would be interesting to see. These investigations would also include finding out what kind of processing delay is acceptable from an end-user perspective.

Chapter 6

Conclusion

The thesis aimed at investigating whether 2D Human Pose Estimation could be used for analysing the technique of a swing in golf. The goals stated in Section 1.3 were all achieved. Technical features that could be detected were successfully identified. These consisted in knee flexion and primary spine angle during the address, and lateral side bending angle at the top of the swing. The data was collected and analysed with the developed systems, which consisted in an application allowing to record a video of the swing and algorithms to identify the phases of the swing and measure joint angles and the lateral bending of the torso. Based on the used methodology, it can be concluded that 2D Human Pose Estimation running on a mobile phone can be used to identify the phases of the swing, knee and hip angles, and lateral bending angle. Additionally, it showed that there is no significant difference in the identification of the events between the sagittal and frontal plane, and that the same algorithm can be used for it. This research highlighted the potential of using 2D Human Pose Estimation in swing analysis. Some issues remain to be solved, but with additional research and development into the features that can be identified in the golf swing, developing an application to give feedback to the golfer using 2D Human Pose Estimation is possible.

6.1 Main findings

The developed phase identification algorithm reached the same performance as an existing open-source phase identification machine learning model. The algorithm identified the toe-up, the top and the impact, which allowed to split the swing into its main four phases with an accuracy of 39% and a mean absolute error of less than two frames. Within fives frames of the actual event, 95% of the events were detected. This all happened independently of the filming angle, in the sagittal or frontal plane, using the same algorithm based on the vertical movement of the wrists throughout the swing. However, the video processing pipeline showed some data loss and creation, as the frames of the processed video and the original video did not always match. This would indicate that the used validation method based on frame numbers was not the best way, and using timestamps could provide a more exact evaluation. The angle measurement algorithms performed well for the knee flexion measurement, being correct in 66% of the videos, and for the lateral side bending, correct in 87% of the videos. The primary spine angle measurement had an offset of about 10 degrees, which made the Percentage of Correct Angles of 0.04. However, the standard deviation of both the error and absolute error was not much bigger than for the knee flexion, which makes it reasonable to assume that with an offset correction or by correcting the systematic error causing the drift, the primary spine algorithm would reach accuracies similar to the knee flexion detection algorithm. This systematic error is likely due to the joint centres not being assessed in the same way in the video and by the pose detection model. Since the offset is only an issue in one of the detected angles, the assessment of the shoulder joint seems to be the issue.

6.2 Remaining open questions

For future works in the area, the main remaining open issue is the frame discrepancy. This includes why do some processed videos have more or less frames than the original frames and how to account for those discrepancies when validating the event detection algorithm. If the used methodology were to be replicated, instead of developing the system to work in real-time from the beginning, the system should be built to allow only post-processing of the data. The first step should thus be to have a working system allowing to validate the data with all the processing happening after the video has been recorded. This would shift the focus to the accuracy and precision of the algorithms, without concerns of data loss due to real-time processing of the pose. Once the algorithms are validated, the causes and consequences on accuracy and precision of real-time processing should be elucidated, and a real-time system developed. As this thesis consisted of an exploratory work into investigating the possibilities and potential of 2D Human Pose Estimation for swing analysis, many aspects are left to dive into. Some most promising ones are to continue the algorithmic approach to identify more aspects of the swing. Using the joint coordinates for machine learning is left unexplored in this thesis but could certainly provide an interesting alternative approach to the original phase identification problem.

References

[1] National Golf Foundation. *Participation and engagement rise again.* 2022. URL: https://www.ngf.org/participation-and-engagement-rise-again/#:~:text=NGF's%5C%20participation%5C%20research%5C%20shows%5C%20overall,on%5C%2D%5C%20and%5C%20off%5C%2Dcourse. (visited on 01/17/2023).

[2] Adrian M. Burden, Paul N. Grimshaw, and Eric S. Wallace. "Hip and shoulder rotations during the golf swing of sub-10 handicap players". In: *Journal of Sports Sciences* 16.2 (1998). PMID: 9531005, pp. 165–176. DOI: 10.1080/026404198366876.

[3] X. Zhang and G. Shan. "Where do golf driver swings go wrong? Factors influencing driver swing consistency". In: *Scandinavian Journal of Medicine & Science in Sports* 24.5 (2014), pp. 749–757. DOI: https://doi.org/10.1111/sms.12061. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/sms.12061.

[4] Nathan Edwards, Clark Dickin, and Henry Wang. "Low back pain and golf: A review of biomechanical risk factors". In: *Sports Medicine and Health Science* 2.1 (2020), pp. 10–18. ISSN: 2666-3376. DOI: https://doi.org/10.1016/j.smhs.2020.03.002. URL: https://www.sciencedirect.com/science/article/pii/S2666337620300068.

[5] Jo Armour Smith, Andrew Hawkins, Marybeth Grant-Beuttler, Richard Beuttler, and Szu-Ping Lee. "Risk Factors Associated With Low Back Pain in Golfers: A Systematic Review and Meta-analysis". In: *Sports Health* 10.6 (2018). PMID: 30130164, pp. 538–546. DOI: 10.1177/1941738118795425.

[6] Adam P Cann, Anthony A Vandervoort, and David M Lindsay. "Optimizing the benefits versus risks of golf participation by older people". In: *Journal of geriatric physical therapy* 28.3 (2005), pp. 85–92. DOI: 10.1519/00139143-200512000-00004.

[7] Louis Passfield and James G. Hopker. "A Mine of Information: Can Sports Analytics Provide Wisdom From Your Data?" In: *International Journal of Sports Physiology and Performance* 12.7 (2017), pp. 851–855. DOI: 10.1123/ijspp.2016-0644. URL: https://journals.humankinetics.com/view/journals/ijspp/12/7/article-p851.xml.

[8] Aritz Badiola-Bengoa and Amaia Mendez-Zorzilla. "A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise". In: *Sensors* 21.18 (2002). ISSN: 1424-8220. DOI: https://doi.org/10.3390/s21185996. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8472911/.

[9] Chen-Chieh Liao, Dong-Hyun Hwang, and Hideki Koike. "How Can I Swing Like Pro?: Golf Swing Analysis Tool for Self Training". In: *SIGGRAPH Asia 2021 Posters.* SA '21 Posters. Tokyo, Japan: Association for Computing Machinery, 2021. ISBN: 9781450386876. DOI: 10.1145/3476124.3488645.

[10] Maxime Bourgain, Philippe Rouch, Olivier Rouillon, Patricia Thoreux, and Christophe Sauret. "Golf Swing Biomechanics: A Systematic Review and Methodological Recommendations for Kinematics". In: *Sports* 10.6 (2022). ISSN: 2075-4663. DOI: 10.3390/sports10060091. URL: https://www.mdpi.com/2075-4663/10/6/91.

[11] Sung Eun Kim, Jangyun Lee, Sae Yong Lee, Hae-Dong Lee, Jae Kun Shim, and Sung-Cheol Lee. "Small changes in ball position at address cause a chain effect in golf swing". In: *Scientific Reports* 11.1 (2021), p. 2694. DOI: 10.1038/s41598-020-79091-7.

[12] *Golf technique discussion with professional and amateur golfers.* Transcript available on request. Jan. 2023.

[13] Patria A Hume, Justin Keogh, and Duncan Reid. "The role of biomechanics in maximising distance and accuracy of golf shots". In: *Sports medicine* 35 (2005), pp. 429–449. DOI: 10.2165/00007256-200535050-00005. URL: https://www.researchgate.net/publication/7845632_The_Role_of_Biomechanics_in_Maximising_Distance_and_Accuracy_of_Golf_Shots.

[14] Phillip Page and Todd S Ellenbecker. *The scientific and clinical application of elastic resistance*. Human Kinetics, 2003, pp. 222–223. URL: https://books.google.se/books?id=Q8K8FZyTdw8C&pg=PA222&lpg=PA222&dq=primary+spine+angle&source=bl&ots=KxjyeuBcJG&sig=ACfU3U0hLLOX-dhT4JfNX4FlNegcZurBYQ&hl=en&sa=X&ved=2ahUKEwjRi-zh74L9AhXGR_EDHaF3CEYQ6AF6BAhHEAM#v=onepage&q=primary%5C%20spine%5C%20angle&f=false.

[15] Koji Murakami, Satoshi Hamai, Ken Okazaki, Satoru Ikebe, Takeshi Shimoto, Daisuke Hara, Hideki Mizu-uchi, Hidehiko Higaki, and Yukihide Iwamoto. "In vivo kinematics of healthy male knees during squat and golf swing using image-matching techniques". In: *The Knee* 23.2 (2016), pp. 221–226. ISSN: 0968-0160. DOI: https://doi.org/10.1016/j.knee.2015.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0968016015001738.

[16] Sol-Bi Kim, Joshua (Sung) H. You, Oh-Yun Kwon, and Chung-Hwi Yi. "Lumbopelvic Kinematic Characteristics of Golfers With Limited Hip Rotation". In: *The American Journal of Sports Medicine* 43.1 (2015). PMID: 25398245, pp. 113–120. DOI: 10.1177/0363546514555698.

[17] David M Lindsay and Anthony A Vandervoort. "Golf-related low back pain: a review of causative factors and prevention strategies". In: *Asian journal of sports medicine* 5.4 (2014). DOI: https://doi.org/10.5812/asjsm.24289.

[18] Joseph Myers, Scott Lephart, Yung-Shen Tsai, Timothy Sell, James Smoliga, and John Jolly. "The role of upper torso and pelvis rotation in driving performance during the golf swing". In: *Journal of sports sciences* 26 (Feb. 2008), pp. 181–8. DOI: 10.1080/02640410701373543. URL: https://www.researchgate.net/publication/5987291_The_role_of_upper_torso_and_pelvis_rotation_in_driving_performance_during_the_golf_swing.

[19] Phillip J. Cheetham, Philip E. Martin, Robert Earl Mottram, and Bryan F. St. Laurent. "The importance of stretching the "X-Factor" in the downswing of golf: The "X-Factor Stretch"". In: *Book of abstracts 2000 pre-olympic congress. International congress on sports science medicine and physical education*. 2000. URL: https://www.semanticscholar.org/paper/The-importance-of-stretching-the-%5C%22X-Factor%5C%22-in-the-Cheetham-Martin/f5280146566794f9e1d95d58903ddf8e1eac78e6.

[20] John Hellström. "Competitive Elite Golf A Review of the Relationships between Playing Results, Technique and Physique". In: *Sports medicine (Auckland, N.Z.)* 39 (Feb. 2009), pp. 723–41. DOI: 10.2165/11315200-000000000-00000.

[21] Kazuaki Kawashima, Tetsuo Meshizuka, and Shunichi Takeshita. "A kinematic analysis of foot force exerted on the soles during the golf swing among skilled and unskilled golfers". In: *Science and golf III: Proceedings of the world scientific congress of golf*. Human Kinetics Champaign, Illinois. 1998, pp. 40–45. URL: https://www.tib.eu/de/suchen/id/BLCP%5C%3ACN028918786.

[22] Kevin Ball and Russell Best. "Centre of pressure patterns in the golf swing: individual-based analysis". In: *Sports Biomechanics* 11.2 (2012). PMID: 22900399, pp. 175–189. DOI: 10.1080/14763141.2012.673007.

[23] Theodore T. Kim, Mohamed A. Zohdy, and Michael P. Barker. "Applying Pose Estimation to Predict Amateur Golf Swing Performance Using Edge Processing". In: *IEEE Access* 8 (2020), pp. 143769–143776. DOI: 10.1109/ACCESS.2020.3014186.

[24] A.J. McHardy and Henry Pollard. "Muscle activity during the golf swing * Commentary". In: *British journal of sports medicine* 39 (Dec. 2005), 799–804, discussion 799. DOI: 10.1136/bjsm.2005.020271.

[25] William McNally, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. "GolfDB: A Video Database for Golf Swing Sequencing". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2019.

[26] John Horton and David Lindsay. "Comparison of spine motion in elite golfers with and without low back pain". In: *Journal of Sports Sciences* 20 (Sept. 2002). DOI: 10.1080/026404102320183158.

[27] Yungchien Chu, Timothy C. Sell, and Scott M. Lephart. "The relationship between biomechanical variables and driving performance during the golf swing". In: *Journal of Sports Sciences* 28.11 (2010). PMID: 20845215, pp. 1251–1259. DOI: 10.1080/02640414.2010.507249.

[28] Paul Grimshaw, Andrew Giles, Richard Tong, and Karen Grimmer. "Lower Back and Elbow Injuries in Golf". In: *Sports Medicine* 32.10 (2002), pp. 655–666. ISSN: 1179-2035. DOI: https://doi.org/10.2165/00007256-200232100-00004. URL: https://link.springer.com/article/10.2165/00007256-200232100-00004.

[29] Andrew McHardy and Henry Pollard. "Lower back pain in golfers: a review of the literature". In: *Journal of Chiropractic Medicine* 4.3 (2005), pp. 135–143. ISSN: 1556-3707. DOI: https://doi.org/10.1016/S0899-3467(07)60122-0. URL: https://www.sciencedirect.com/science/article/pii/S0899346707601220.

[30] Eoghan Murray, Emma Birley, Richard Twycross-Lewis, and Dylan Morrissey. "The relationship between hip rotation range of movement and low back pain prevalence in amateur golfers: an observational study". In: *Physical Therapy in Sport* 10.4 (2009), pp. 131–135.

[31] Loren Schwarz, Diana Mateus, and Nassir Navab. "Discriminative Human Full-Body Pose Estimation from Wearable Inertial Sensor Data". In: Nov. 2009, pp. 159–172. ISBN: 978-3-642-10468-8. DOI: 10.1007/978-3-642-10470-1_14.

[32] Mukhiddin Toshpulatov, Wookey Lee, Suan Lee, and Arousha Haghighian Roudsari. "Human pose, hand and mesh estimation using deep learning: a survey". In: *The Journal of Supercomputing* 78 (Apr. 2022). DOI: 10.1007/s11227-021-04184-7.

[33] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. "2D Human Pose Estimation: New Benchmark and State of the Art Analysis". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 3686–3693. DOI: 10.1109/CVPR.2014.471.

[34] Soonchan Park, Ju Yong Chang, Hyuk Jeong, Jae-Ho Lee, and Ji-Young Park. "Accurate and Efficient 3D Human Pose Estimation Algorithm Using Single Depth Images for Pose Analysis in Golf". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 105–113. DOI: 10.1109/CVPRW.2017.19.

[35] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. "Deep 3D human pose estimation: A review". In: *Computer Vision and Image Understanding* 210 (2021), p. 103225. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2021.103225. URL: https://www.sciencedirect.com/science/article/pii/S1077314221000692.

[36] M. Bourgain, S. Hybois, P. Thoreux, O. Rouillon, P. Rouch, and C. Sauret. "Effect of shoulder model complexity in upper-body kinematics analysis of the golf swing". In: *Journal of Biomechanics* 75 (2018), pp. 154–158. ISSN: 0021-9290. DOI: https://doi.org/10.1016/j.jbiomech.2018.04.025. URL: https://www.sciencedirect.com/science/article/pii/S0021929018302987.

[37] Kerrie Evans, Sean A. Horan, Robert J. Neal, Rod S. Barrett, and Peter M. Mills. "Repeatability of three-dimensional thorax and pelvis kinematics in the golf swing measured using a field-based motion capture system". In: *Sports Biomechanics* 11.2 (2012). PMID: 22900406, pp. 262–272. DOI: 10.1080/14763141.2012.654502.

[38] Antanas Verikas, Evaldas Vaiciukynas, Adas Gelzinis, James Parker, and M. Charlotte Olsson. "Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness". In: *Sensors* 16.4 (2016). ISSN: 1424-8220. DOI: 10.3390/s16040592. URL: https://www.mdpi.com/1424-8220/16/4/592.

[39] Zhongyu Jiang, Haorui Ji, Samuel Menaker, and Jenq-Neng Hwang. "GolfPose: Golf Swing Analyses with a Monocular Camera Based Human Pose Estimation". In: *2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. 2022, pp. 1–6. DOI: 10.1109/ICMEW56448.2022.9859415.

[40] Trackman. *Trackman 4: technology.* 2023. URL: https://www.trackman.com/golf/trackman-4/tech-specs (visited on 01/23/2023).

[41] Toptracer. *Toptracer range.* 2023. URL: https://toptracer.com/range/ (visited on 01/23/2023).

[42] Zepp. *Train with Zepp.* 2022. URL: http://www.zepplabs.com/en-us/golf/smart-coach/ (visited on 01/19/2023).

[43] Blast. *Blast Golf.* 2022. URL: https://golf.blastmotion.com/golf/ (visited on 01/19/2023).

[44] Garmin. *Garmin TruSwing.* 2022. URL: https://www.garmin.com/en-US/p/505083 (visited on 01/19/2023).

[45] Kazunari Takeichi, Masaru Ichikawa, Ryota Shinayama, and Takehiro Tagawa. "A Mobile Application for Running Form Analysis Based On Pose Estimation Technique". In: *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (2018), pp. 1–4. DOI: 10.1109/ICMEW.2018.8551559.

[46] OnForm. *Video analysis app and coaching platform for instant feedback. Save time, coach smarter and stay organized.* 2022. URL: https://www.onform.com/ (visited on 01/17/2023).

[47] Android Developers. *CameraX overview.* 2023. URL: https://developer.android.com/training/camerax (visited on 01/20/2023).

[48] Google ML kit. *Pose detetction.* 2023. URL: https://developers.google.com/ml-kit/vision/pose-detection?hl=en (visited on 01/20/2023).

[49] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. "BlazePose: On-device Real-time Body Pose tracking". In: arXiv, 2020. DOI: 10.48550/ARXIV.2006.10204. URL: https://arxiv.org/abs/2006.10204.

[50] Valentin Bazarevsky, Ivan Grishenko, and Eduard Gabriel Bazavan. *Model card: MediaPipe BlazePose GHUM 3D.* Google. June 2021. URL: https://developers.google.com/static/ml-kit/images/vision/pose-detection/pose_model_card.pdf.

[51] Ritesh Kanjee. *Top 9 Pose Estimation Models of 2022.* 2022. URL: https://medium.com/augmented-startups/top-9-pose-estimation-models-of-2022-70d00b11db43 (visited on 01/18/2023).

[52] Elisha Odemakinde. *Human Pose Estimation with Deep Learning - Ultimate Overview in 2022.* 2022. URL: https://viso.ai/deep-learning/pose-estimation-ultimate-overview/ (visited on 01/18/2023).

[53] BeomJun Jo and SeongKi Kim. "Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices". In: *Traitement du Signal* 39.1 (2022), pp. 119–124. DOI: https://doi.org/10.18280/ts.390111. URL: https://www.iieta.org/journals/ts/paper/10.18280/ts.390111.

[54] Sarah Mroz, Natalie Baddour, Connor McGuirk, Pascale Juneau, Albert Tu, Kevin Cheung, and Edward Lemaire. "Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose". In: *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART).* 2021, pp. 1–4. DOI: 10.1109/BioSMART54244.2021.9677850.

[55] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, Bill Freeman, Rahul Sukthankar, and Cristian Sminchisescu. "GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Oral).* 2020, pp. 6184–6193. URL: https://openaccess.thecvf.com/content_CVPR_2020/html/Xu_GHUM__GHUML_Generative_3D_Human_Shape_and_Articulated_Pose_CVPR_2020_paper.html.

[56] J. Guiñón, Emma Ortega, José García-Antón, and V. Pérez-Herranz. "Moving average and Savitzki-Golay smoothing filters using Mathcad". In: (Jan. 2007). URL: https://www.researchgate.net/publication/228407245_Moving_average_and_Savitzki-Golay_smoothing_filters_using_Mathcad.

Appendix

A Participants

Gender	Lead side	Handicap	Years of experience
Woman	left	6.5	25
Woman	left	7	18
Woman	left	36	>20
Woman	left	38	5
Woman	left	N/A	0
Woman	left	N/A	0
Man	left	0	30
Man	left	5.7	>20
Man	left	9	35
Man	left	9.4	10
Man	left	20	10
Man	left	20	20
Man	left	26	4
Man	left	N/A	0
Man	left	N/A	0

Table A.1: Detailed description of the participants

B Results of the event identification

Plane	Event	PCE0	PCE1	PCE2	PCE5
Sagittal & frontal plane	Toe-up	0.41	0.51	0.72	0.96
	Top	0.32	0.56	0.74	0.95
	Impact	0.38	0.56	0.76	0.95
	All	0.37	0.55	0.74	0.95
Sagittal plane	Toe-up	0.38	0.51	0.70	0.96
	Top	0.32	0.62	0.77	0.94
	Impact	0.38	0.51	0.77	0.94
	All	0.36	0.55	0.74	0.94
Frontal plane	Toe-up	0.45	0.52	0.74	0.97
	Top	0.32	0.48	0.71	0.97
	Impact	0.39	0.65	0.74	0.97
	All	0.39	0.55	0.73	0.97

Table B.2: Percentage of correctly detected events

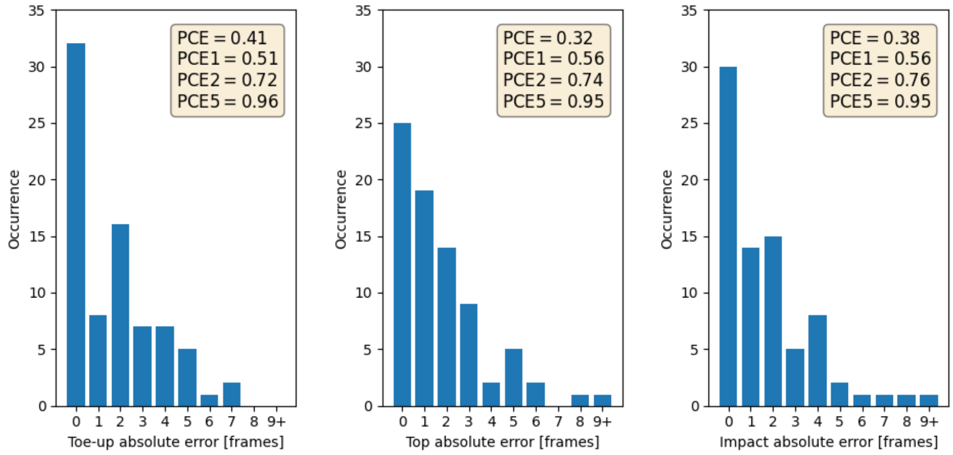


Figure B.1: Absolute error distribution for each event, with their PCE metrics

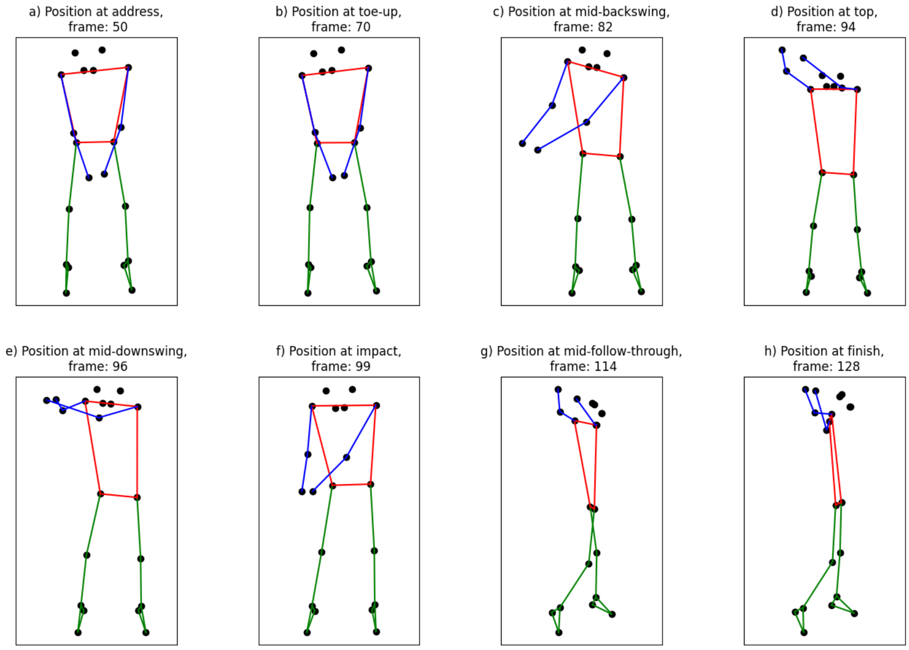
Plane	**Event**	**Mean absolute error**	**Standard deviation**	**Mean error**	**Standard deviation**
Sagittal & frontal plane	Toe-up	1.72	1.89	0.41	2.53
	Top	1.74	1.95	0.23	2.61
	Impact	1.69	2.05	0.54	2.61
	All	1.72	1.96	-	-
Sagittal plane	Toe-up	1.79	1.91	-	-
	Top	1.66	1.90	-	-
	Impact	1.77	2.01	-	-
	All	1.74	1.93	-	-
Frontal plane	Toe-up	1.61	1.89	-	-
	Top	1.87	2.05	-	-
	Impact	1.58	2.13	-	-
	All	1.69	2.01	-	-

Table B.3: Absolute errors and errors in frames for event detection, with their standard deviation

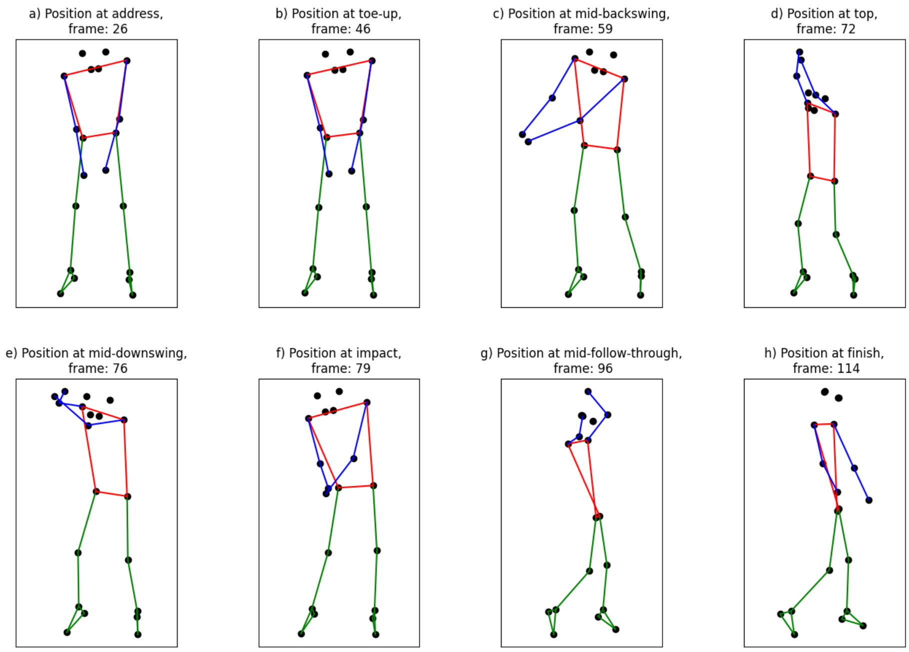
Event	**PCE**	**PCE1**	**PCE2**	**PCE5**	**Mean absolute error in frames**
Toe-up	0.48	0.52	0.76	0.95	1.67 ± 2.08
Top	0.32	0.62	0.62	0.95	2.05 ± 2.16
Impact	0.33	0.52	0.71	0.95	1.76 ± 1.89
All	0.38	0.51	0.70	0.95	1.83 ± 2.02

Table B.4: Complete results by event for the event frames predicted by the SwingNet model

C Skeleton plots



(a) Example 1



(b) Example 2

Figure C.2: Two skeleton plots of the golfer throughout the swing, in the frontal plane

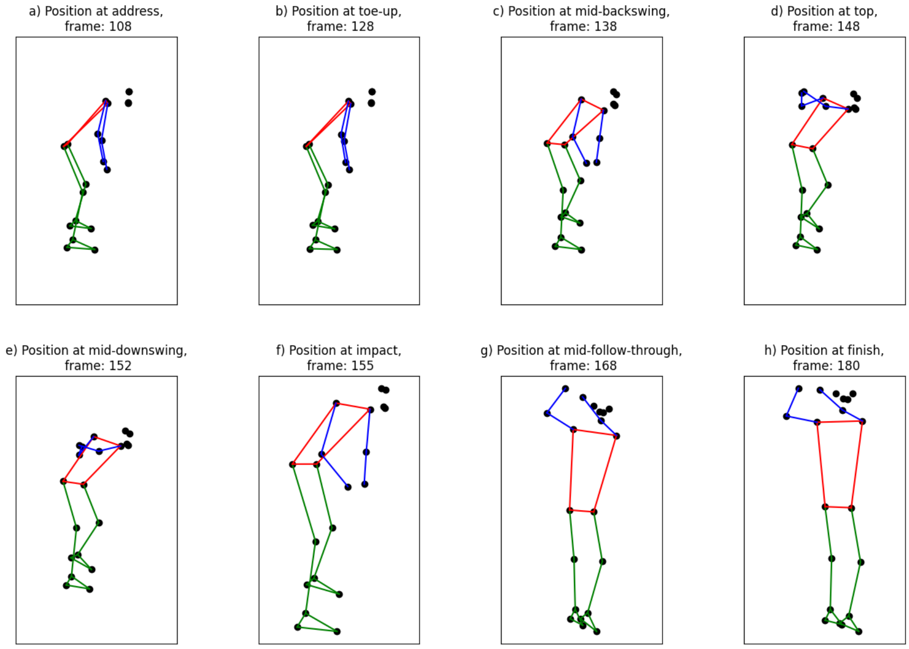


Figure C.3: Skeleton plot of the golfer throughout the swing, in the sagittal plane

TRITA – CBH-GRU-2023:057