Towards using artificial intelligence as tool in artistic gymnastics coaching – case backward giant circle

Master's thesis
University of Turku
Department of Computing
Master's degree in artificial intelligence
2021
Petteri Laitinen

Supervisors:
Petra Virjonen
Jukka Heikkonen

UNIVERSITY OF TURKU
Department of Computing

PETTERI LAITINEN: Towards using artificial intelligence as tool in artistic gymnastics coaching
– case backward giant circle

Master's thesis, 47 p., 4 app. p.
Master's degree in artificial intelligence
May 2021

The objective of this thesis was to study whether it is possible to create a system that estimates artistic gymnast's body joint angles based on a low-budget 2-dimensional single RGB video recording. To meet the objective, 54 video files were collected on gymnasts performing backward giant circle skill, together with assessments of the performances by two professional coaches. The video files contained total of 233 repetitions of the skill. A pilot system of computer vision algorithms was developed, using an open source human body pose recognition algorithm. An algorithm based on pixel grayscale value was developed and used to recognize starting and ending moment of a repetition and to sample each repetition at 7 key phases. Body joint angle estimates were calculated based on the body part location estimates of the 1631 samples. The work proved that it is possible to develop a system that estimates body joint angles of an artistic gymnast. It was found that rotation and cropping of the frames improved probability of yielding correct estimates. The angle estimate for knees had highest, up to 66%, correlation with coach evaluations. Hips and shoulders had weak but significant correlation with coach evaluations. The results indicate that it is possible to develop a low-budget system that could work as augmented tool in artistic gymnastics coaching. In addition, human body pose recognition provides a new method to biomechanical research of artistic gymnastics.

Keywords: computer vision, artistic gymnastics, backward giant circle, uneven bars, human body pose estimation, body joint angle estimation, markerless pose estimation

Contents

1 Introduction **1**

2 Literature **6**

2.1 Development of human body pose detection as analytics enabling methodology in sports 6

2.2 Machine learning and computer vision in sports 9

2.3 Machine learning and computer vision in artistic gymnastics 10

2.4 Studies on backward giant circle 12

3 Research questions **14**

4 Materials and methods **15**

4.1 Data 15

4.1.1 Video files 15

4.1.2 Coach evaluations 16

4.2 Composition of pilot system that estimates gymnast's body joint angles . 18

4.2.1 Phase detection based on pixel grayscale value 18

4.2.2 Recognition of number of giants performed 19

4.2.3 Sampling, rotation and cropping of key frames 23

4.2.4 Body joint angle estimation 25

4.2.5 Pilot feature extraction algorithm for body joint angle data 26

4.3 Verification of validity of the angle estimates 28
4.3.1 Verification through visual inspection 28
4.3.2 Verification through post-conditions 29
4.4 Correlation between angle estimates and coach evaluations 29

5 Results 32
5.1 Body joint angle estimating system for artistic gymnastics 32
5.1.1 Phase detection based on pixel grayscale value 33
5.1.2 Recognition of number of giants performed 33
5.2 Validity of the body joint angle estimates 33
5.2.1 Body joint angle estimates based on the original frames 33
5.2.2 Body joint angle estimates based on the rotated and cropped frames 34
5.3 Angle thresholds that maximize correlation with coach evaluations 36

6 Discussion and conclusions 38
6.1 Interpretation of the results 38
6.1.1 Body joint angle estimating system for artistic gymnastics 38
6.1.2 Sensibility of the body joint angle estimates 40
6.1.3 Correlation between angle estimates and coach evaluation 41
6.2 Practical implications 43
6.3 Limitations 44
6.4 Possible future work 45

References 47

Appendices

A Example pictures A-1

B OpenPose body25 model and angle definitions B-1

List of Figures

1.1 The nine phases of the backward giant circle. 4

4.1 Camera positioning and centering 16

4.2 Four types of performances 21

4.3 Performance types as time series 22

4.4 Sample collection phases 23

4.5 Angle trigonometrics 26

4.6 Feature extraction algorithm as flowchart 27

5.1 Average joint angle estimates 35

5.2 Correlation between selected angle estimates and coach evaluations . . . 37

A.1 Video scene example A-1

A.2 Example frame with up-and-down misestimate A-1

A.3 Rotated and cropped example frame A-2

B.1 Body25 model of OpenPose B-1

List of Tables

5.1 Visual inspection results of original frames 33
5.2 Visual inspection results of rotated frames 34
5.3 Angle estimate acceptance rates per camera 34
5.4 Angle estimate acceptance rates at different phases. 35
5.5 Descriptive statistics of angle estimates 36
5.6 Optimal angle thresholds 37

B.1 Body points used when calculating the joint angles. B-2

Abbreviations and notions

The below list describes several abbreviations and key notions that will be later used within the body of the document

Abbreviations

AI Artificial intelligence

CoP Code of Points - rulebook of women's artistic gymnastics owned by International Gymnastics Federation

FIG International Gymnastics Federation

FPS Frames per second, unit of frame rate

M Arithmetic mean value

MAG Men's artistic gymnastics

N Number of observations

r Pearson correlation coefficient

RGB Format of color video, where each pixel gets a value of red (R), green (G) and blue (B)

STD Standard deviation

WAG Women's artistic gymnastics

Notions

Annotation

Explanation or comment added next to a picture, part of it or other object.

Artificial intelligence Behavior of a machine that is so humanlike that it is considered as intelligent. The machine can be limited to a very narrow problem, and when it can (usually through supervised or unsupervised learning) solve the problem in a humanlike way, its behavior fulfills the definition of artificial intelligence.

Backward giant circle

Element on women's uneven bars, where gymnast rotates 360 degrees around the higher bar with face towards the movement direction in regular grip. Element number 3.201 in Code of Points.

Body joint angle

Angle defined by a joint and its neighboring joints. A fully extended joint has an angle of 0 degrees. If the joint is ahead of the line between the two neighboring joints in the direction of the backward giant circle movement, the angle is defined as negative. If the joint is behind the line between two neighboring joints, the angle is defined as positive.

Coach evaluation

To facilitate this study, two coaches had evaluated each body joint at 9 phases on a 0/1 scale, 0 corresponding to pure pose at the joint, 1 corresponding to fault.

Computer vision

Field of computer science that works on enabling computers to derive meaningful information from visual input such as picture or video. Computer vision is also considered as subdiscipline of artificial intelligence.

Element

A gymnastics element is a single move in a routine that has a value. The elements are listed in FIG's Code of Points.

Frame

Still picture from a video file. A video file consists of consequtive frames.

Performance

Demonstration of a skill by gymnast that can be judged. A performance during a training can include several consequtive repetitions.

Phase

Notation used in this study to describe the progress of backward giant circle element. Defined as clockwise angle between the gymnast and upwards pointing vertical line from upper bar, measured in degrees.

Skill

Synonym to element. A single move in artistic gymnastics that has value in a competition.

Uneven bars

One of the four events in women's artistic gymnastics. Also called asymmetric(al) parallell bars.

1 Introduction

I was introduced to artistic gymnastics when my daughters started it as a hobby somewhere around year 2012. I still remember the feeling when I got to visit the training hall for the first time. I instantly fell in love with the atmosphere: the dust of chalk in the air, the aspiration for high performance, the never-ending persistence of gymnasts trying a new skill over and over again, and the combination of beauty, power and flexibility in successful performances that I thought was impossible. This thesis is in a way my tribute to all the gymnasts. My daughters have now in 2021 quit the artistic gymnastics, but my admiration for artistic gymnastics still remains.

Women's artistic gymnastics (WAG) is a sports competition discipline, where athletes perform elements on four apparatus: vault, uneven bars, balance beam and floor exercise. The performances shall demonstrate agility, artistry, flexibility, power and style, and are judged according to rules governed by the International Gymnastics Federation (FIG). [1]

At the uneven bars apparatus, gymnasts perform swinging elements around two asymmetric bars (upper bar on 2.5 m height and lower bar on 1.7 m height), frequently releasing the bar and recatching it [1]. *Backward giant circle* is a key skill that enables gymnasts to proceed to release-regrasp skills like Tkatchev and various dismounts [2]. During the backward giant circle, the gymnast rotates a full 360 degree circle around the upper bar with body extended from handstand to handstand. She has palms facing backwards and face directed to the direction of rotation. The backward giant circle element is coded as element 3.201 in WAG Code of Points [3].

The apparatuses of women's artistic gymnastics (WAG) and men's artistic gymnastics (MAG) are different. Backward giant circle is also performed on high bar in MAG, but mainly due to the absence of the low bar, the way of performing the backward giant circle is somewhat different between high bar and uneven bars.

My subjective observation as an occasional visitor of the gymnastics training and competition facilities was that gymnastics coaches use quite little video recording as tool in coaching. The coach was often next to the gymnast assisting, and had no possibility to do video recording. On the other hand, the training took place repetitively - the exactly same skill was trained in the exactly same place over and over again. The thought arised: could I develop a tool for coaches that helped them giving instant feedback to the gymnasts about their performances? This thesis documents the practical learnings from the efforts of trying to create such a tool using methods of computer vision, feature extraction and machine learning.

The work is limited to one gymnastics skill: backward giant circle on uneven bars. The backward giant circle was selected as the studied element after interviewing local gymnastics coaches. According to the coaches, the giant circle is a very rapid movement, and despite being visually easy movement, it is complicated to explain in words. Hence it felt like a very good candidate for computer vision and machine learning. The coaches also concluded that an augmented coaching tool would be welcome.

To facilitate the study, data was collected in autumn 2019 and winter 2020 by RGB video recording young gymnasts performing the backward giant circle. The preparations for collecting data, the procedures of video recording on-site and the resulted data are described in detail in a separate technical report [4].

The research done on the backward giant circle has so far mostly had a solid basis on traditional biomechanics, where simulations of four-segment spring-mass systems have been pruned and verified using data from video recordings, for example in the work by Yeadon and Hiley [5]. The present work had more of a data-driven approach - the video data was collected first, and the videos were thereafter used to build tailored computer vision and machine learning algorithms. According to Simmons and Chappell [6], the term artificial intelligence (AI) denotes behavior of a machine which, if a human behaves in the same way, is considered intelligent. The original ambition was to build a system of algorithms that would mimic coach behaviour. The system would take in a video as input and return a human-like evaluation of the recorded performance, therefore fulfilling the previous definition of AI.

No AI algorithm makes sense stand-alone. AI can be seen as a vehicle to solve a problem, but to solve a problem, there should be enough knowledge about the context. Sometimes increasing the knowledge leads to re-definition of the problem, which in turn leads to a different solution than originally anticipated. Gymnastics coaches are the experts that probably possess most knowledge about artistic gymnastics. To gain more knowledge about the context, two very professional coaches were interviewed. They were asked to assess the video recorded performances. Based on the coaches' suggestion, each repetition was assessed in 9 different phases, and the purity of elbows, shoulders, hips, knees and toes/leg split was separately assessed on a 0/1 scale, 0 meaning pure and 1 meaning fault [4]. The phases are depicted in the Figure 1.1. A good AI algorithm would probably mimic coach assessments. The coach assessment seemed to be very much based on experience-based knowledge about the angles of elbows, shoulders, hips, knees and toes at each phase. Therefore a system that would estimate the body joint angles seemed to be a good target. During unformal interviews, gymnasts stated they wish one or absolutely maximum two points of corrective feedback, in order to have proper focus on the corrections. Trying to change too many parts of a skill includes a danger of losing control, that in turn leads to increased risk of severe injury! According to the gymnasts, the feedback should also be easy to understand: specific and tangible, something that speaks for feedback in form of 'straighten your elbows at handstand', or 'flex your hips more when you pass the lower bar' - to give some examples. This kind of feedback can be given based on

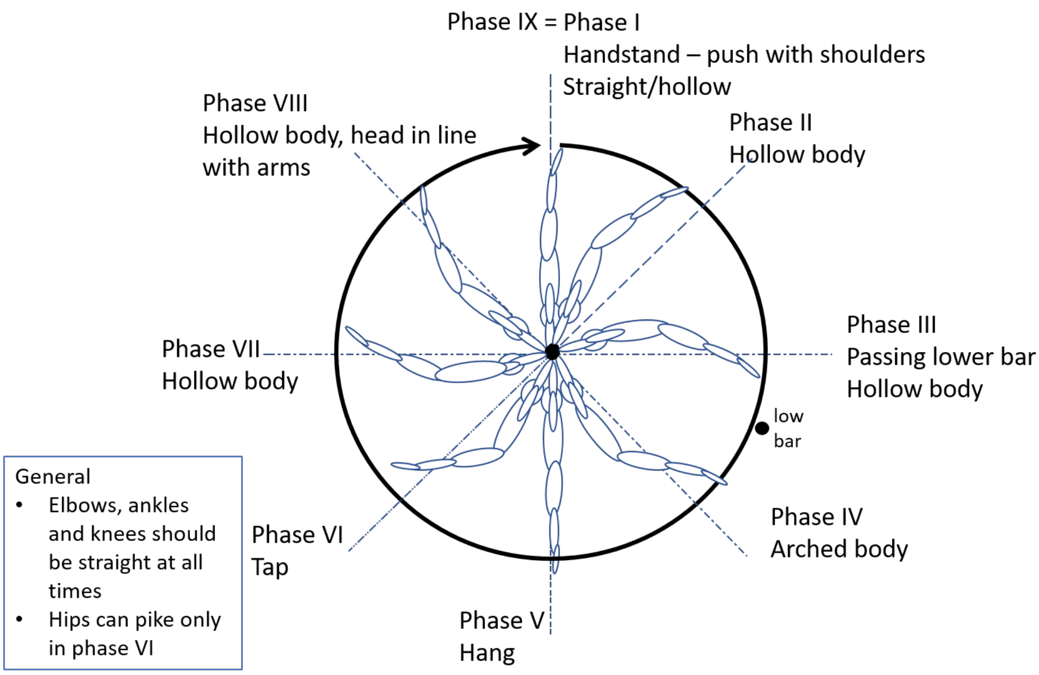


Figure 1.1: The nine phases of the backward giant circle.

body joint angle estimates.

In order for a tool to be meaningful for a coach, it should be easy to use, as the coach has the gymnast and the gymnast's safety in focus. Therefore, it should be enough that the coach or an assistant just starts and stops the video recording. The tool should then give the feedback in based on recorded video. This means that the software should recognize the number of repetitions of the skill. It should also recognise when each repetition starts, and when it ends.

Once the starting and ending times of repetitions are known, feature extraction methods can be applied to extract data that describes the performance. Background subtraction techniques cover methods to detect moving objects in a static scene, and could come useful in the setup. In this study, a person keypoint-detection method called Openpose [7], was used, to extract coordinates of gymnast body parts. Using the coordinates, joint angles could be calculated, that in turn could be categorised or used directly as input to a feedback prioritization algorithm.

Gymnastics is reported to be a hard source of image data for person keypoint-detection algorithms. A typical pose of the gymnast has two elements that make it hard. First, it is normal that gymnasts rotate 360 degrees, so the body can be at any angle. Second, gymnasts keep arms and legs tightly together or next to body, making parts of the body less visible. The poses are rare in normal life, and consequently there are less pictures available to train a keypoint-detection algorithm. Further improving of the used keypoint-detection method was left outside of the scope of this study, as it would have required extensive annotation of frames from videos. The results of the visual inspections of body pose and angle estimates given in Section 5.2 give some new insight how big of a problem the rotation movement of the gymnast is, and what happens if the frames are rotated so that gymnast appears always to be ‘hanging‘ upright prior to feeding the frame into body pose recognition algorithm.

The Chapter 2 starts by a short summary of key studies in human body pose recognition. Thereafter, a brief walk-through of latest research is given, where machine learning and computer vision in particular is applied in sports, narrowing gradually down to artistic gymnastics and studies on the backward giant circle in particular. Thereafter, the objective of this study, broken down into three specific research questions, is stated in the Chapter 3. The Chapter 4 describes the methods and data to answer the research questions. It focuses mainly on the development work of the computer vision related algorithms, as the data and the collection of the data is described separately in a technical report [4]. The results are reported in the Chapter 5 and discussed in the Chapter 6, mirroring them against the three research questions. The thesis ends with some self-criticism and suggestions for possible future work.

One of the possible paths of future work is to build a full feedback system - this thesis shows that it is possible to extract body joint angle estimates from a single RGB video, but the practical application is left to future!

2 Literature

The work in this thesis combines computer vision and machine learning with sports. The scientific research field is therefore very broad. The literature review aims at giving a short a review of one selected area of computer vision: human body pose estimation. After that, a brief review on computer vision in sports in general is given, followed by a review of research in artistic gymnastics. Finally, a review of the research of the backward giant circle element is presented.

2.1 Development of human body pose detection as analytics enabling methodology in sports

It is trivial for a human to recognize location of a person in a normal picture. Human can easily calculate number of people in the picture, and point the exact positions of body parts of each person in the picture. For a computer, the same picture is just a matrix of color values [8]. Human body pose detection algorithms are computing applications that try to recognize persons and their body parts in a picture or a video frame [9]. These algorithms are typically trained using big databases with images that have been manually labelled. They can be applied as such, or if needed, they can be further trained by adding more pictures with manual labels, or by changing the architecture of the models.

With big data, there is an emerging market for models that recognize and understand human gestures and body posture, leading to open source projects. The leading open source project at 2020 that can be applied on video or picture is OpenPose, with application for multi-person detection including body, face and hand-detection [10]. The fundamentals of the realtime multi-person 2D pose estimation model of OpenPose are described in [7], and further improvements on face and hand recognition of multiple persons in the image are described in [11]. The face and hand recognition were outside of the scope of this study.

The multi-person pose estimation models build on single-person pose detection models. Single-person pose detection models base on assumption that there is only one person of interest in the image, with somewhat known location. Gong et al. categorize the single-pose models into three categories: kinematic, planar and volumetric models [9]. The kinematic models are of main interest here. They typically consist of sticks corresponding to parts of skeleton. The ends of the sticks correspond to body joints, and the model usually has a tree-structure. The models can have additional rules to capture occlusion, body symmetry and long-range relationships [7], [9]. Kinematic models can be based on a pre-defined model, or alternatively, the model can be built by a graph structure learning from image data [9].

One example of methodology used to improve the pose detection methods is the concept of poselet. Bourdev introduced the term of poselet in 2009 [12], by suggesting that it describes a part of one's pose. To use poselets requires a database with annotated poselets, which also was a part of Bourdev's work in 2009. According to Bourdev, a good poselet fulfils two criteria. First, it should be easy to find the poselet given the input image. Second, it should be easy to localize the 3D configuration of the person conditioned on the detection of a poselet. This leads to a good poselet being as tightly clustered part of image as possible. Poselets have thereafter been applied for example by Bourdev [13], Pischulin [14], and Hernández-Vela [15] to improve the posture modeling methods. There are other, alternative methods that have been used to improve the posture modeling methods [9].

Standardized and annotated data is needed to build models, but also to measure goodness and performance of the developed models. One example of annotated database is MPII Human Pose dataset [16] that includes around 25 thousand images of everyday human activities with annotated body joints, collected from Youtube, and is available at [17]. Another database is Leeds Sports Pose Extended Training Dataset [18] available at [19] that contains 10 thousand images gathered from Flickr searches for the tags 'parkour', 'gymnastics', and 'athletics' and consists of poses deemed to be challenging to estimate.

These databases can be used to measure model performance together with standardized model evaluation criteria. There are three common evaluation criteria used in the field:

- PCP: Probability of a correct pose that measures the percentage of correctly localized body parts. A candidate body part is labeled as correct if its segment endpoints lie within 50% of the length of the ground-truth annotated endpoints. [20]
- PCK: Probability of a correct keypoint [21]. Works like PCP but with tighter bounding of the keypoints.
- APK: Average precision of keypoints [21]. APK calculates average of correctly classified body parts over all body parts in the picture, and requires all the persons in the image to be modeled.

All these measures require annotated images (bounding boxes) to work. Both PCK and APK define a candidate keypoint to be correct if it falls within $a \cdot \max(h, w)$ pixels of the ground-truth keypoint, where h and w are the height and width of the bounding box respectively, and a controls the relative threshold for considering correctness.

The state-of-the-art methods today base more or less all on convolutional neural networks. Newell et al. showed in 2016 that convolutional neural networks in so called stacked hourglass form work efficiently in single-person pose detection modeling [22]. They also stress that a degree of meta-knowledge is essential in real-use systems to tackle situations like occluded body parts. Usually, occlusion is due to three reasons: 1) part of the body is behind another object 2) part of the body is outside the image borders 3) part of the body is covered by another part of the body. The occlusions and special scenes have been reported to remain as main challenge for human pose models by Hua et al. in 2020 [23].

Action recognition using convolutional neural networks on top of combination of RGB and depth video has been studied using different data sets. The study concluded that by combining multiple parallell models using both RGB and depth data is key to reach a robust classification model for human action [24].

Less is reported about sensibility, validity and reliability of the body joint angle estimates based on kinematic human body pose models.

2.2 Machine learning and computer vision in sports

The use of machine learning and computer vision in sports is ever-evolving, and the field is developing in rapid speed, making it difficult to have a complete overview. An overview of the sport movement classification was given in 2019 that listed 52 sport-specific studies ranging from team ball sports like volleyball, football and soccer to individual sports like swimming, diving and karate [25]. The only listed study on gymnastics in that review was the study of Diaz-Pereira et al. that built a concept algorithm that recognizes selected movements from a video record of a rhythmic gymnastics performance and predicts the judge scoring [26]. There are sport movement studies on artistic gymnastics, too, like the work carried out by Mack et al. [27]. A deeper review of studies focusing on computer vision in artistic gymnastics is given in Section 2.3.

The field is manifold, and it is typical that machine learning and computer vision are tools that help finding a solution to relatively narrow problem. The basic methods can be transferred from one discipline to another one with a similar problem, but require a high level of adaption to fit into the new area of sports. A tool to provide competition swimmers with fast feedback was introduced by Nevalainen et al. in 2016, with data collected using several under-water cameras and processed by methods available in open source libraries, allowing a low cost budget solution [28]. Using convolutional neural network models to estimate the pose of a swimmer has also been used, and it has been shown to be possible to increase the pose recognition by annotation of frames from videos [29]. Nibali et al. have used selected frames from a set of systematically collected videos on diving, and trained temporal action localization neural network to estimate the action in the frame [30]. The outcome is an estimate of what kind of a dive the athlete performs in the video. Kanth et al. presented a new application in 2018 where they applied computing on data from a single camera to determine the step length, speed, and the feet-contact-time of a pole vault athlete [31].

2.3 Machine learning and computer vision in artistic gymnastics

The use of machine learning and computer vision in sports is increasing, and during the years 2018–2021 there are also reported studies that have applied these methods in artistic gymnastics.

Artistic gymnastics is a competitive sports discipline where performance is measured by subjective judgements, i.e. judges observe the performance and evaluate it using criteria defined in a handbook called code of points [3]. There is a natural aspiration for as objective judgement as possible. Hence, one branch of usage of computer vision and machine learning in artistic gymnastics is related to judging.

Mack et al. (2019) assumed that gymnastics skill performances with similar kinematic patterns over time lead to similar evaluation scores, and that similarity can be observed through monitoring the main body angles throughout the skill performance. They recorded gymnasts' performances using a digital video camera with 240 frames per second and 1920 x 1080 pixels resolution, placing the camera with its optical axis orthogonal to gymnast's movement direction, simulating the judge's perspective. They developed four models (one recurrent neural network model, and three variants of nearest-neighbour) for floor, balance beam and vault skills and their compared predictions against original scores using Spearman's rank correlation. The results indicated that recurrent neural network had more promising prediction ability than the nearest-neighbour variants [27].

Fujitsu announced in November 2019 that it has built a commercial application targeted to artistic gymnastics, basing on an 18-body-point posture model and multiple laser emitting sensors [32]. The system was for the first time officially used in an FIG competition as an additional tool for confirming difficulty scores on four apparatus in late 2019: men's pommel horse, men's still rings, and men's and women's vault [33]. Augmented information about gymnast performance will most likely be provided to spectators of the coming Tokyo Olympic Games through a further developed version of this tool as part of Fujitsu's ambitions to popularize the 3D sensing technologies.

In order to facilitate machine learning, a database of annotated videos or pictures of performances is required. Competition performances provide a natural source of data to this kind of database when combined with the evaluation scores. However, there is also room for coaching-oriented annotation of performances. The annotations can describe generic body postures (piked/straddle/hollow/straight/overextended) or individual joints (toes pointed/toes flexed) or rotation degree. Very little is known about usability of such an approach. The present work contributes to creating understanding of using systematic coach evaluations of body parts during execution of an element.

Summing up, the focus of applying computer vision in artistic gymnastics seems to be to secure fair judging. Less is reported about computer vision as coaching tool. Video based feedback can potentially be a pedagogical tool in fostering motor learning, motivation and self-assessment during physical education programme with young children [34]. Fujitsu believes that providing judges with AI tools will create demand for AI tools in coaching and training [32]. For gymnasts to score high in a competition they need to train in a way that corresponds to judging principles. If judging is based on AI, natural way to meet the AI judging is to use AI in training. Hence, it seems important to study computer vision and AI as coaching tool. On the other hand, it has been stated that artistic gymnastics coaches can benefit from a greater understanding of the mechanisms that control gymnasts' skill development [35], [36]. Therefore computer vision and AI will provide value in coaching if it can support knowledge creation.

2.4 Studies on backward giant circle

The research of elements on uneven bars has solid basis in biomechanics. The basic kinetics of backward giant is well-described by Witten in 1990 [37]. In the work, the author used combination of strain gage, force transducer and video camera recording to study forces and moments that athletes expose to during the element. However, the distance between the bars has increased in the past 30 years, and consequently also the performance of the giant element has somewhat changed since 1990. Today, the lower bar shall be minimum 130 cm and maximum 180 cm from the upper bar on uneven bars [1]. Consequently a female gymnast must straddle, pike or arch to pass the lower bar. It has been found that video recording and analysis can provide with information to coaches and researchers to select gymnast-specific technique to pass the lower bar during the backward giant circle prior to a Tkatchev dismount [36]. Several camera set-ups with and without markers have been tried out by Manning in her doctoral dissertation [2]. One difficulty especially in competition situations is to find a suitable positioning of the cameras. In competition situations the camera has to be placed so that a reprojection of the video using camera calibration and planar rectification is needed.

Also men's horizontal bar apparatus has the element backward giant, but it is performed slightly differently, reason being twofold: absence of low bar, and different radius

(4,0 cm in WAG vs 2,8 cm in MAG) and friction coefficient of the bar itself (uneven bars have natural wooden surface, horizontal bar of polished steel). Hence, the results of the studies on men's horizontal bar should not be directly applied on women's uneven bars. However, the same methods should be applicable. Biomechanics of the backward giant on men's horizontal bar has been studied using video recording by Yeadon and Hiley in the 21st century [38], focusing on giant prior to dismount [5], triple piked somersault dismount [39] and Tkatchev [40]. Novices' learning of the longswing (very close to backward giant circle) element has been studied using video recordings [41]. Active markers placed on the body of the athlete were usually used in the studies using video recording [37],[42],[41],[40], but this is not always possible, like in competitions, where more complex calculation using data from two (or more) cameras is needed [5], [43].

Overall, majority of the previous work has relied on biomechanical spring-mass models and the video observations have been handled manually to measure body posture in order to prune and verify the parameters of biomechanical models. Apart from Fujitsu's commercial program to apply laser based depth-sensing technology using several sensors [32] and an 18-body-point posture model, the area of applying human body pose estimating models on artistic gymnastics seems unexplored so far. The fact that biomechanics of backward giant circle is pretty well-known and well-described makes it as a fruitful element to test human body pose models and possibility of explaining backward giant circle performance through body joint angle information.

3 Research questions

This study aims to increase understanding of what it takes to create a low-budget AI tool that could help artistic gymnastics coach to assess the backward giant circle skill on uneven bars, and work as a virtual second pair of eyes.

Generally speaking, a skill is time-dependent series of actions that the gymnast performs. The actions start at the time point when the gymnast starts the skill, and end at the time point when the gymnast has performed the skill. On the other hand, a skill is also position-dependent series of actions that the gymnast performs. The body point locations over time define the skill and the performance. The assessment of the performance is based on body poses at different phases of the performance. A body pose, in turn, can be defined through the angles of body joints.

The objective of this study is to find out whether it is possible through applying computer vision methods to create a system of algorithms that provides meaningful feedback about the gymnast's performance. On-site implementation is left out of scope.

The main research questions, limited to the backward giant circle on uneven bars, are:

1. Is it possible to create a system that estimates gymnast's body joint angles based on low-budget 2-dimensional single RGB video recording?
2. If such a system can be created, in which circumstances and to what extent does the system provide with sensible body joint angle estimates?
3. How do the body point angle estimates produced by the system correlate with coach evaluation?

4 Materials and methods

The objective of the present study was to find out whether it is possible to create a low-budget system with a single RGB video camera that records gymnast's performance and gives tangible, specific and meaningful feedback about the performance. There are four different apparatuses in women's artistic gymnastics, and numerous different skills that gymnasts could demonstrate on each of them. After talking to the local gymnastics coaches, the scope was limited to one basic skill: backward giant circle on uneven bars.

4.1 Data

4.1.1 Video files

To facilitate the study, a data set was collected during autumn 2019 and early 2020 using video recording [4]. The recordings produced of 54 videos with a total of 291 backward giant circle elements performed by 11 gymnasts. An illustrative video scene example is given in the Appendix in the Figure A.1, where four frames of one performance are blended on top of each other. The video camera used in recordings was mainly a single Osmo Pocket, model OT110, a commercially available low-budget camera that has remote controlling possibility from mobile phone, features self-leveling, and takes high density video of 1920 x 1080 pixels with 60 frames per second. Ten videos were recorded with Canon EOS 1300D and EFS 18-55mm objective, using frame width of 1280 pixels, frame height 720 pixels and 50 frames per second. Both cameras use H.264 (also known as

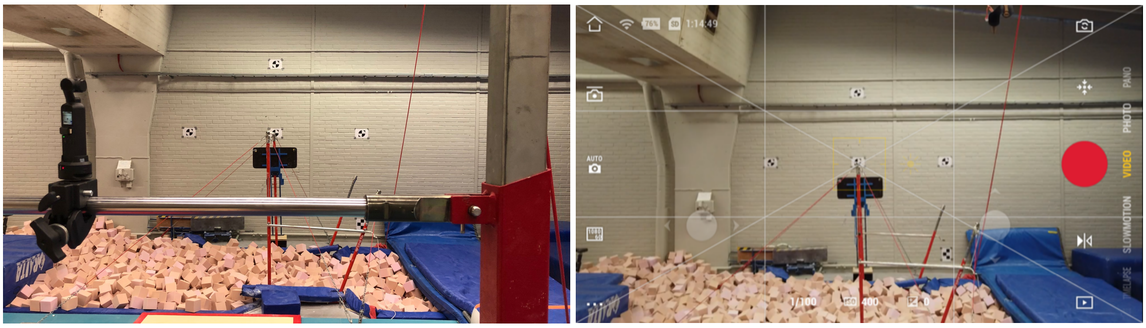


Figure 4.1: Camera positioning (left) and camera centering (right)

MPEG-4 Part 10/AVC) video compression standard [44], [45].

The camera was placed so that the gymnast was fully inside the captured video scene during the entire skill. Also, the supporting board was moved to the other side of the apparatus, so that only the narrow support tube frames of the bars were inbetween the gymnast and the camera. As the camera was placed in line with the bar in order to take the pictures in sagittal plane, a very special type of occlusion happened. In perfect performances, the gymnast had so symmetric pose that you could see only the left side of the body in the picture. This was the cost of using a single camera and placing the camera perpendicular to the sagittal plane so that a reprojection of the video using camera calibration and planar rectification was not needed. The placement of the camera in line with the upper bar and centering using the diagonal grid of the camera steering software are depicted in the Figure 4.1.

4.1.2 Coach evaluations

Machine learning can be split into two disciplines: unsupervised and supervised learning [46, p. 4]. In unsupervised learning, the algorithms try to find and describe patterns in the provided data without a specified target variable. In supervised learning, a target variable is provided, and the algorithms try to describe the target variable as well as possible.

The recorded video files would have been sufficient to facilitate unsupervised learning approach, and if a target variable could be possible to extract from video files, they could facilitate even supervised approach. However, reflecting towards the ambition to create a tool that is intelligent and work as a human-being would, it was decided to ask professional coaches to evaluate the video recorded performances.

Two professional gymnastics coaches, both undergone FIG Academy for coaches and holding FIG's judge Brevet [47] that certifies to act as judge in international competitions [47] (in fact, both of them have acted as judge in the Olympic Games), were asked to evaluate the performances. The evaluation was agreed to have a coaching perspective rather than a judge perspective. Further, the evaluation was agreed to follow the technique preparing for a turning element, like pirouette. Backward giant circle is used also to prepare for a release move like Tkatchev and for dismount off the bars. Both of these have a bit different technique than the backward giant circle preparing for a turning element, that starts from a handstand and ends clearly at a handstand.

The evaluations resulted in one row of data for each repetition of the movement. Each row contained video ID, repetition number and 45 binary variables indicating whether there was any fault in the given body part at the given phase. Body parts assessed were elbow, shoulder, hip, knee and ankle area. Phases were at 45 degrees intervals starting from handstand (0 degrees phase) and ending at handstand (360 degrees phase), leading to 9 phases illustrated in the Figure 1.1. As there were 291 repetitions with five body parts analyzed at 9 phases, there were 13 095 data points recorded into the database. The procedures of creating the data are reported in a technical report of Turku Centre For Computer Science [4]. The ankle area assessments included observation whether the legs were split in an uncontrolled manner. Due to this ambiguity, the ankle/leg split observations were removed from the data before further analysis.

4.2 Composition of pilot system that estimates gymnast's body joint angles

4.2.1 Phase detection based on pixel grayscale value

In very general terms, a digital video file consists of frames that are consequtive images at consequtive time points. A frame consists of a matrix of pixels, and the number of vertical vs horizontal pixels gives the height and width of the frame and video. A pixel has a color value, which can be represented in several ways. In the current study, the pixels (and hence videos) had 8-bit RGB representation. This means that a pixel has three values, one for red (R), one for green (G) and one for blue (B) color, ranging from 0 to $2^8 - 1 = 255$. Black color has the representation $[0, 0, 0]$ and white $[255, 255, 255]$.

It is often practical to have just one value per pixel instead of three. In this study, the RGB values were transformed to grayscale by convex combination of the red, green and blue value:

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B, \tag{4.1}$$

rounding to the nearest integer value [48, p. 15]. There are variety of different ways to do the color conversions. Testing different color conversions and their effect to the quality of computer vision algorithms were left out of the scope of this study.

In object recognition, a frame is divided into two parts: background and foreground. Background subtraction methods aim to reduce the background to focus to the interesting part [49], [50]. In the current study, the gymnast is the foreground, or the object to be recognized. Rest of the scene (including the parts of the apparatus that are between the gymnast and the camera) are background. The body of gymnast covers roughly a sector of 20 degrees of the total 360 degrees full circle around the upper bar, meaning that the gymnast covers roughly 5-6% of the "directions" from the upper bar. When the gymnast is in full movement, it takes roughly 2 seconds (1,5 to 2,5 s) for her to circle around the bar. With a frame rate of 60 frames per second, there are 120 (90 to 150) frames captured during one circle. This means that if we only study a certain pixel, the gymnasts body covers the pixel on average during 7 consecutive frames, and during 113 consecutive frames the pixel represents color value of a static background. The angular velocity of the gymnast is highest in the swinging phase at the bottom of the giant circle, where the odds ratio between gymnast and the background is even smaller than 7/113, and during the handstand phase, the odds ratio is easily double the normal, 14/106.

This indicates that by studying the time series of grayscale values of a pixel we can make a fairly good estimate of whether the gymnast is in the current angle. Let $Y(p,t)$ be the grayscale value of pixel p with coordinates (p_x, p_y) at time point t. Then the grayscale values $\{Y(p,t)|t = 0, 1, ..., N-1\}$, where N is number of frames in the video, can be assumed to follow a normal distribution around a value that corresponds to the background. Whenever the gymnast passes the pixel, there is a change in the pixel color value, that can be recognized by setting a threshold S:

$$G(t) = \begin{cases} 1 & \text{if } \left\|\frac{Y(p,t)-\bar{Y}}{\sigma_Y}\right\| > S \\ 0 & \text{otherwise,} \end{cases} \tag{4.2}$$

where σ_Y is the standard deviation of grayscale value Y and $\bar{Y}$ is the arithmetic mean value of Y. Theoretically, this is outlier detection, and in general, value $S = 3$ is a good starting point [51, p. 554], as 99,7% of normally distributed data should fall within the borders. Experiments with different values showed, however, that threshold value $S = 2.4$ provided best results for the current data, so that was used throughout the study.

4.2.2 Recognition of number of giants performed

A backward giant circle starts from a handstand and ends at a handstand. At handstand, gymnast is at 0 degree phase. Judges look at the whole body of gymnast. If any part of the body is more than 30 degrees from the vertical line at the start or at the end, the movement is not counted as giant circle. If any part of the body is more than 10 degrees from the vertical line, reductions can be made. The somewhat loose condition for what is counted as a giant circle required a mapping of different types of performances of the backward giant circle, in particular what comes to the first and last repetition during a performance of several repetitions.

In the collected video data, the gymnasts started with swinging back and forth, followed by a kip-cast, where they rotated counter-clockwise to reach the starting handstand. In majority of the performances, the handstand was incomplete, and the first giant circle consequently started from a phase of 10 – 30 degrees. In some few performances, the starting handstand after kip-cast was complete with gymnast's body pointing at 12 o'clock direction.

The last giant circle could either end in a continued clockwise rotation after the very last handstand, gymnast typically either flexing their body to reduce the angular momentum or continuing to a release from around 250 degrees phase, or it could end in gymnast shifting to counter-clockwise rotation, which was typically the case if the gymnast did not reach a complete handstand.

The combination of two alternatives for the first handstand and two alternatives for the last handstand give us four different scenarios depicted in the Figure 4.2. The four scenarios can theoretically be caught by studying pixel color values of two points, a and b, that should be on equal distance from the upper bar and at suitable phase. The points should be on such distance from the bar that the silhouette of the gymnast (including hair!) never overlaps the pixels in any other movement than giant circle. Typical example of such a case would be when the gymnast is leaning on the bar with straight elbows - this is called front support in terms of artistic gymnastics. On the other hand, the distance from the bar should be minimum possible so that even short gymnasts pass the pixel while circling, and because the cartesian velocity of the body is slower on shorter distances, leading to higher accuracy of the signal. Time series patterns of the standardized pixel color values

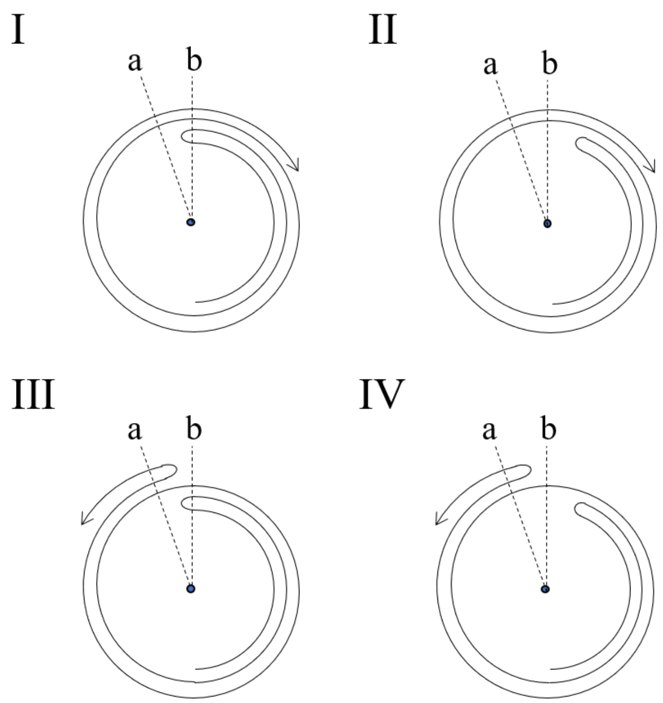


Figure 4.2: Four types of performances illustrated. I: complete kip-cast to handstand in the start and backward rotation after last handstand, II: incomplete kip-cast plus backward rotation at end, III: Complete kip-cast but incomplete handstand at end of series, IV: Incomplete kip-cast in the start and incomplete handstand at the end.

corresponding to the scenarios of the Figure 4.2 are depicted in the Figure 4.3. Giants starting from a complete handstand are boxed with dash-dotted line, giants ending to counter-clockwise rotation with dotted line and regular giants with dashed line. This is a rough simplification of the patterns. For example, if the rotation direction changes right on top of a pixel, there is no double peak, but just one.

The exact position of the points a and b depends also on the background. In the current study, the background was a brickwall painted in white. If possible, the neighbourhood of the background of the pixel positions should be as homogenous in color value as possible. This could be arranged by covering potential interest areas with a plain white or black canvas or paper. Black background should work well if the pixel positions are placed so that bare skin of legs cover it while gymnast passes, but less well if the gymnast wears black tights, which was the case in some of the videos in the current study. Even the harm-

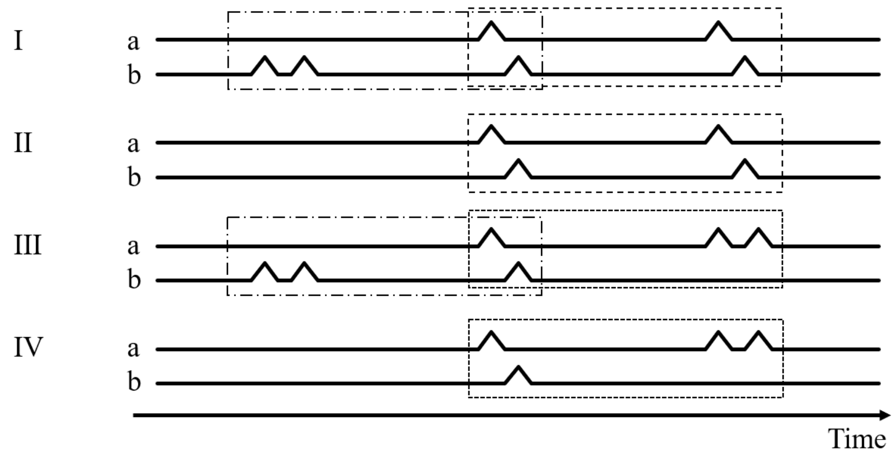


Figure 4.3: Illustration of time series of pixel values of the four scenarios.

less looking brickwall texture caused challenges to the positioning of the pixel points, as the texture had somewhat different color values and the camera was placed on 10 meters distance from the wall, leading to a small vibration in the camera fixture to cause slight shift of the picture. On the other side, the lighting of the indoor hall was very stable, thanks to the fact that there were no windows and no daylight variations present.

A simple program was written to recognize the frames where the gymnast passed the 12 o'clock position in a clockwise rotation. The recognition was based on studying the grayscale value of a pixel using the Equation 4.2. Based on this type of simple rule, a "regular" backward giant circle was recognized as interval between two consequtive events of passing the 12 o'clock position, as depicted in the Figure 4.3 with dashed box. The points a and b were at -20 degrees phase and 5 degrees phase, respectively, and at distance from the upper bar equal to 140/480 of the distance between the 9 o'clock marker and 3 o'clock marker (measured in number of pixels in the video). A relative measure to the distance of the two markers was used in order to make the method camera-independent. This distance corresponds to the hip and thigh area of gymnasts when the body is fully extended. The code is stored and available in the resulting repository [52].

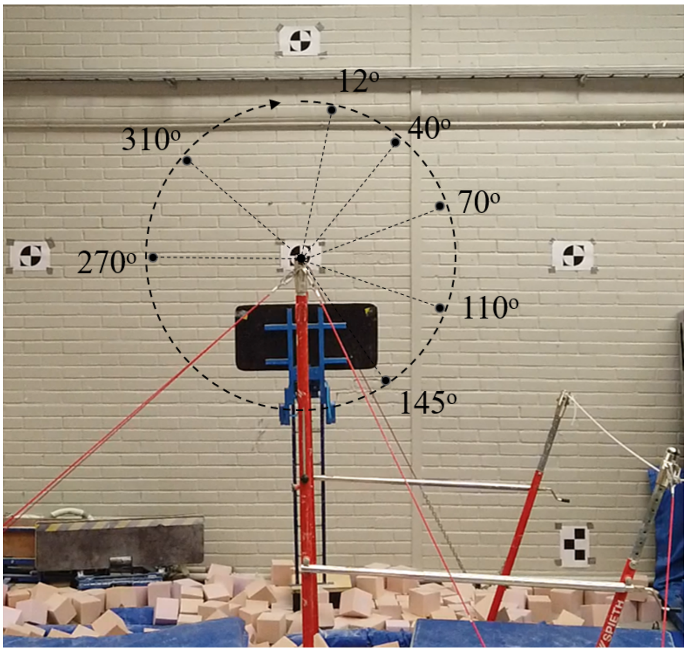


Figure 4.4: Sample collection phases. Body part coordinate estimates were collected at 12, 40, 70, 110, 145, 270 and 310 degrees angles.

4.2.3 Sampling, rotation and cropping of key frames

For each recognized giant circle, data was collected at seven phases: when the front of the body passed 12, 40, 70, 110, 145, 270 and 310 degrees, as depicted in the Figure 4.4. Data was collected more frequently during early phases, primarily based on the coaches' feedback that the latter half of the giant circle is more of an outcome of what has happened during the first half of the giant circle. The coaches saw it important for the execution of the whole giant circle, that the gymnast masters the early phases of the giant circle. Secondarily, the speed of the gymnast is highest around 180 degrees phase leading to motion blur in the frame. Last, the vertical supports and supporting cable structures interfer the computer vision during phases from 150 degrees to 230 degrees.

It would have been possible to collect data from every frame during the performance, or it could have been possible to collect data every 5 or 10 degrees, or even at every degree. It was a decision to try to collect data from points that were as similar to the 9 phases chosen by the coaches (see Figure 1.1). Early pilot tests revealed that body pose recognition algorithm produced misestimates and lacking estimates. Therefore it was decided to scope out efforts to try to describe body part trajectories as time series, but focus on small number of key frames. Discarding the phases of very rapid movement and disturbing support structures led to choosing the points shown in the Figure 4.4. Collecting data from seven angles only was believed to have the benefit of being able to create classification that is possible to communicate to the athlete and coach. 12 degrees corresponds to the moment of starting the backward giant circle from the handstand, 40 degrees corresponds to the phase II of the Figure 1.1 with a hollow body, 110 degrees to the moment of passing the low bar, 145 degrees to the phase IV of the Figure 1.1 with arched body, 270 degrees the phase VII of the Figure 1.1 when going upwards, and finally 310 degrees the phase VIII of the Figure 1.1 when starting to approach the final handstand. The 70 degrees phase did not aim to have a directly matching coach evaluation, it was more an outcome of wish to have more frequent sampling of the first half of the skill.

When the gymnast was recognized to be at certain angle, the frame was rotated by the same angle so that the gymnast was upright, and cropped so that the gymnast fitted into the resulting picture, but majority of the surrounding scene was deleted. This procedure had four benefits. First and most important, having the gymnast upright increased the quality of pose recognition. The early tests showed that sometimes when the gymnast was at handstand, the pose recognition algorithm could misestimate such a case to be an upright pose. An example of such a case is given in Figure A.2 in Appendix A. Second, any other people in the video that could be detected, were effectively removed. Third, narrowing the surroundings decreased chance of detecting false positives from the background. Finally, the cropping was believed to speed up the calculation, even though no exact comparison between algorithms with and without cropping was done. Example of a rotated and cropped frame with a body pose estimate is given in the Appendix A, Figure A.3.

4.2.4 Body joint angle estimation

The sampled frames were fed into a local implementation of a body pose recognition model, to estimate the position of body joints. An open source pose recognition model called OpenPose [7], [11], available at [10], was used. It was decided not to annotate any frames, nor re-train the convolutional neural network that is the core of the OpenPose algorithm. OpenPose was used as it was.

For each person found in a picture, OpenPose returned a matrix that contained body point pixel coordinates and a score between 0 and 1. A score near 1 indicated a strong confidence of the presence of the body point in the given position. Numbering of the body points using OpenPose presentation model is given in the Appendix B, Figure B.1.

Joint angles were calculated from the pixel coordinates of three joints. Let $\mathbf{b} = (b_x, b_y)$ be the coordinates of the pixel corresponding to the joint, angle of which we wish to calculate, and $\mathbf{a} = (a_x, a_y)$ and $\mathbf{c} = (c_x, c_y)$ the coordinates of the adjacent joints that define the angle. We can then calculate the angle θ_b of the joint $\mathbf{b}$ using the trigonometrics sketched in the Figure 4.5:

$$\theta_b = \alpha - \beta = \arctan \frac{b_x - a_x}{b_y - a_y} - \arctan \frac{c_x - b_x}{c_y - b_y}. \tag{4.3}$$

Using this definition, the angle θ measures deviance from full extention in degrees. A fully extended joint has angle of 0 degrees. The Equation 4.3 works well for angles with absolute value below 90 degrees, and gives a negative angle value whenever the joint is in overextension. During a successful backward giant circle performance all the main body joint angles are well below 90 degrees in absolute value.

Angle was calculated for elbow, shoulder, neck, hip, knee, ankle and leg split. Also, angle of the line between wrist and hip and the vertical line from wrist was calculated to check that the phase recognition following the Equation 4.2 had worked as expected. The selection of the points that define the angles is given in Appendix B in the Table B.1.

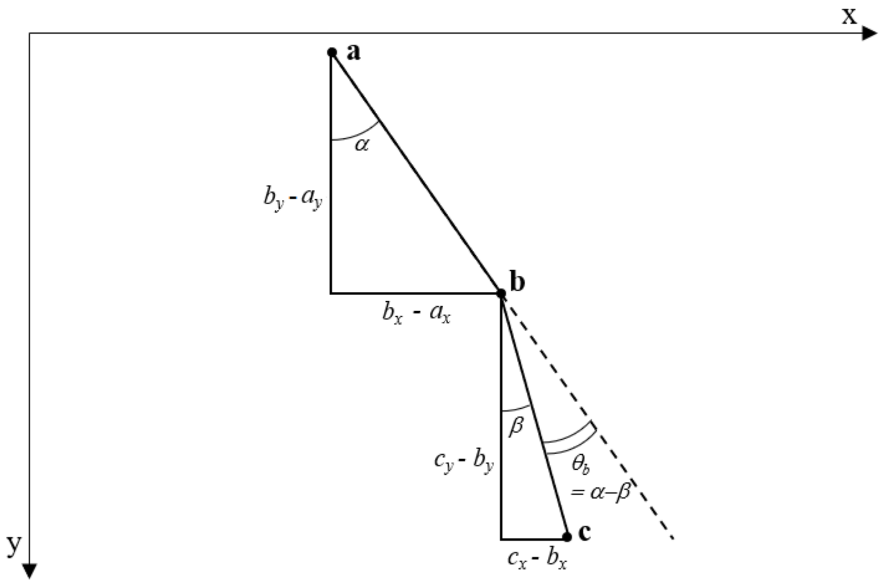


Figure 4.5: Angle trigonometrics. The angle θ_b can be expressed as difference between angles α and β that in turn can be calculated using arc tangent Equation 4.3.

4.2.5 Pilot feature extraction algorithm for body joint angle data

Combining the methods described in Chapters 4.2.1, 4.2.2, 4.2.3 and 4.2.4, a pilot feature extraction algorithm was composed. The pilot feature extraction algorithm is depicted as a flowchart in the Figure 4.6. In the flowchart, there is an additional module for joint correctness classification. Using a threshold, the estimated joint angles can be classified as straight/not to provide binary features. Optimal thresholds can be set based on data analysis between (continuous) angle estimates and binary coach evaluations of the correctness of joint postures. The output of the feature extraction algorithm is then a set of binary joint correctness indicators that can be used as input to an algorithm that based on the indicators ranks potential elements of feedback and points at the most important feedback. Transforming the joint angle estimates into correctness indicators is believed to enable understandable feedback to the gymnasts in form of "Flex your hips more when passing the lower bar", to name one example. The hip correctness indicator at 110 degree phase would then trigger such feedback.

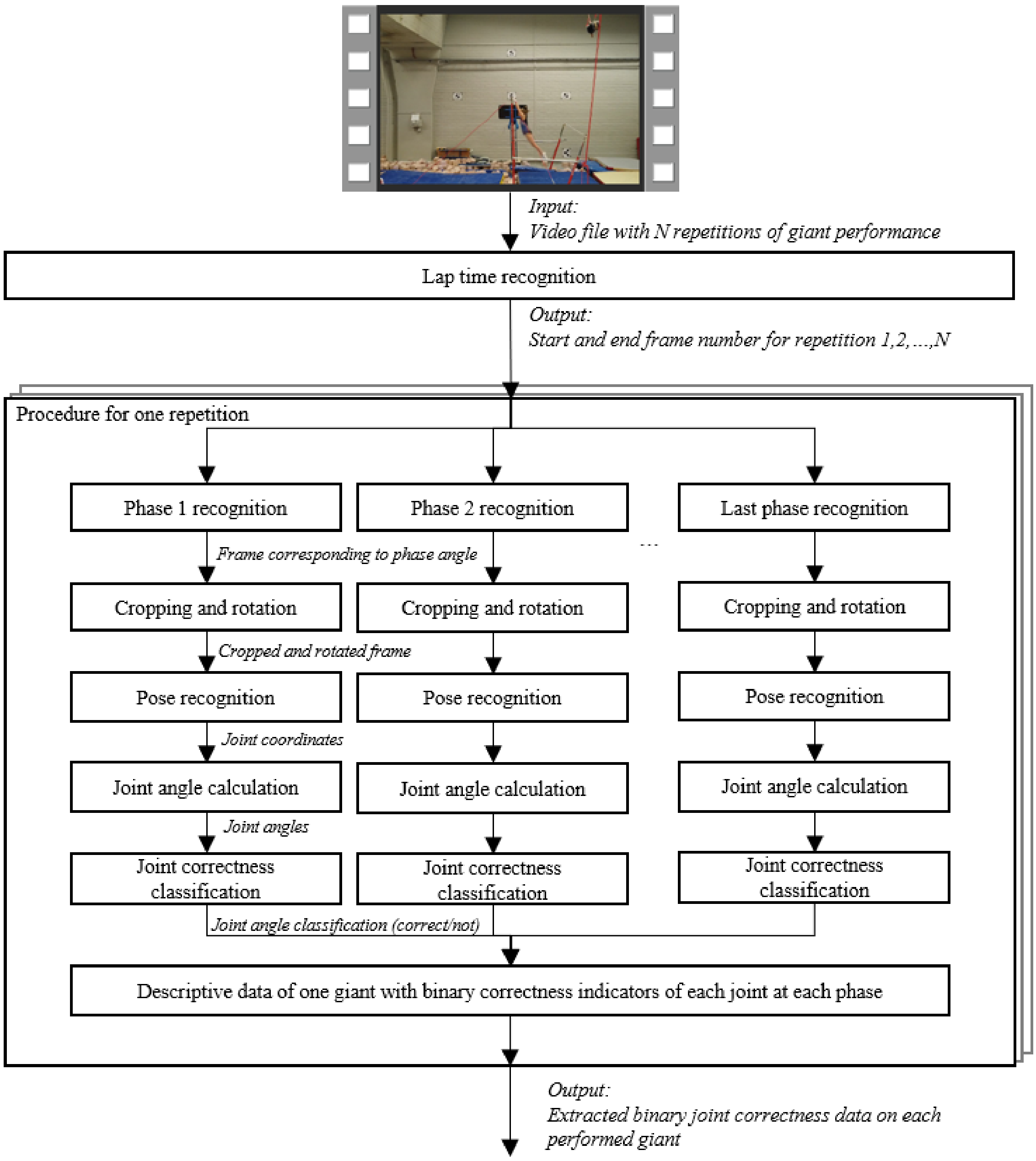


Figure 4.6: Feature extraction algorithm as flowchart

4.3 Verification of validity of the angle estimates

4.3.1 Verification through visual inspection

It is challenging to create a full set of rules to categorize whether the body point estimates and consequent angle estimates are valid. The body pose algorithm could detect all body parts correctly. It could also misestimate some or all parts of the body leading to false angles, it could misinterpret an object in the background as a human, or if there was another outsider person in the image, it could detect the wrong person. Even if the score was over a predefined threshold, there could be misclassification, and on the other side even estimates with low score could in theory still be accurate. Because of all these possibilities, the angle estimates were first inspected visually by going through the images one by one, before building a set of postconditions [53]. Both the original frames and the rotated and cropped (from original frames) pictures were fed into the OpenPose, and the estimated body skeletons were drawn on top of a copy of the original image. Also the angle estimates were printed on the copy of the image. Finally, the copies were stored, and visually inspected. Each picture was categorised into one of the following groups:

- OK estimate (sensible estimates for all angles available)
- Quite OK estimate (sensible estimates for almost all angles available)
- Several angles missing from estimate (sensible estimates for one or two angles)
- No estimate
- False up and down estimate
- Outsider person appearing in the frame
- Something else falsely estimated as person

This was done for 1631 original frames, and separately for 1631 cropped and rotated (from original frames) images. The purpose was to shed light on how sensible the body joint angle estimates are, and discover if there are certain phases where the estimates are systematically lacking. Finally, a comparison between the results for original frames and transformed images gives insight whether cropping and rotation improves the probability of yielding sensible angles.

4.3.2 Verification through post-conditions

After the visual evaluation, it was decided to build a set of post-condition rules that were applied on the rotated and cropped frames. Because the results of the visual evaluation indicated that rotation and cropping improved the sensibility of the angle estimates, it was decided to omit an analysis of post-condition rules on original frames. The post-condition rules were applied to replace an angle by missing value, if the angle was physically unnatural (for example knee angle overextended more than 15 degrees) or not corresponding to nature of backward giant circle (phase-specific conditions for each joint), if two neighboring joints were located too close to each other or too far from each other than viable (then angles of both joints were discarded), and if the y-coordinates of two joints were in wrong order indicating an up-and-down misestimate. A pose or an angle within a pose could be discarded by one or several of these rules, so there was no book-keeping of why an angle was replaced by missing. The acceptance rate of the angles were calculated, defined as the share of angle estimates that passed the post-conditions over all angle estimates.

Finally, descriptive statistics of the angle estimates were produced for each phase.

4.4 Correlation between angle estimates and coach evaluations

The third main research question asked how the angle estimates correlate with coach evaluations. The underlying hypothesis is that there is a correlation between joint angles and gymnast poses that coaches assess. In order to understand the connection between the angles provided by pose recognition and the coach evaluations, Pearson correlation coefficients between joint angles (linear variables, measured in degrees) and coach evaluations (binary variables, 1 corresponding to fault) were first studied. Generally, calculating and trusting on correlation coefficient between a dichotomic and continuous variable is not recommended. In this study, observations with angle estimate close to arithmetic mean value got too small relative weight, whereas observations with angle estimate close to minimum or maximum were overweighted. This also makes Pearson correlation coefficient measure vulnerable to pose recognition providing a false angle estimate (outlier), as on outlier angle estimate could lead to large changes in the correlation as coach recognitions can only take two values.

A potential solution to this challenge is to transform the angle θ to binary variable b using a threshold:

$$b = \begin{cases} 0 & \text{if } \theta \leq \theta_t \\ 1 & \text{otherwise}, \end{cases} \tag{4.4}$$

where θ_t is the threshold angle. This transformation reformulates the question of correlation between angle estimates and coach evaluations: if we were to mimic coach evaluations using only a simple threshold-based rule (*if joint angle is less than X degrees, the posture of the joint is judged as pure*), what would be an optimal threshold? The optimal threshold can be easiest found by calculating correlation coefficient using incremental thresholds to find the angle that maximizes the absolute value of the correlation coefficient. In this approach, only the angle estimate that corresponds to the coach evaluation, is used. For a concrete example: if the coach evaluation of knees at 270 degrees phase is modeled, only the angle estimate of knee at 270 degrees (or closest phase available) is used - no other body parts are used, and no other phases either.

The Pearson correlation coefficients were calculated using different thresholds values between binary transformations of angle estimates and coach evaluations. This was done one by one for elbow, shoulder, hip and knee angle at 12, 40, 110, 145, 270 and 310 degrees of phase. The hypothesis was that 12 degrees angle corresponds to the starting handstand (coaches stressed the starting movement from handstand when evaluating the handstand), 40 degrees angle corresponds to the coach evaluations at 45 degrees phase, 110 degrees angle corresponds to the coach evaluations at 90 degrees phase (coach evaluations stressed the passing of low bar at approximately 110 degrees when evaluating the 90 degrees phase), 145 degrees angle corresponds to the arch phase at 135 degrees in coach evaluations, 270 degrees angle corresponds to the coach evaluations at 270 degrees phase, and finally 310 degrees angle corresponds to the coach evaluations at 315 degrees phase. The differences in angles between coach evaluations and frame sampling were motivated by increased stability in pixel background.

The threshold values that optimize the Pearson correlation between angle and coach evaluation were calculated. There was an underlying hypothesis (based on talking to coaches), that normally a joint could have a wrong angle only in one direction - either too large angle or too small angle. This hypothesis motivated for seeking one global minimum or maximum in the correlation, depending on the phase and joint. The correlation could be positive or negative, depending on joint and phase. Therefore optimization could mean maximization or minimization. A negative correlation meant that if the angle was over the threshold, the joint pose was likely to be assessed as pure by coaches. A positive correlation meant that if the angle was below the threshold, the joint pose was likely to be correct.

To monitor the significance of the correlation coefficients, the classical two-tail test with null hypothesis that correlations coefficient is not equal to zero, was used: $r \neq 0$. Using a significance level of $\alpha = 0.05$ in a t-test, we can derive that the correlation coefficients r are significant if

$$|r| > \frac{e^{1.96\frac{2}{\sqrt{N-3}}} - 1}{e^{1.96\frac{2}{\sqrt{N-3}}} + 1}, \tag{4.5}$$

where N is the number of observations.

5 Results

5.1 Body joint angle estimating system for artistic gymnastics

The research question 1 asked whether it is possible to create a system that estimates gymnast's body joint angles based on low-budget 2-dimensional single RGB video recording. The Chapter 4.2 described a pilot composition of a system of applied computer vision algorithms that based on video data estimates gymnast's body joint angles during performance of a backward giant circle skill on uneven bars. The system used fixed positioning of single RGB camera perpendicular to the movement (described in [4]). The produced video files were pre-processed and sampled. Sampled frames were fed into an open-source body pose algorithm to receive pixel coordinates of body joint positions. The pixel coordinates of body joint positions were finally geometrically transformed to angles that describe the performance. Consequently, it can be stated that it is possible to create such a system that estimates gymnast's body joint angles for the backward giant circle on uneven bars.

Subsections 5.1.1 and 5.1.2 report key success measures of the algorithms for phase detection and recognition of number of giants. Thereafter, the results of angle estimate validity are given in the Section 5.2 and optimal angle thresholds that maximize the correlation with coach evaluations are reported in the Section 5.3.

5.1.1 Phase detection based on pixel grayscale value

The phase detection based on pixel grayscale value described in the Chapter 4.2.1 using the Equation 4.2 with value S=2,4 was verified through visual inspections. In the manual visual inspections, 6 frames out of 1631 were noticed to be sampled at a wrong moment, while rest were sampled at appropriate point of time. The correctness of the phase detection was therefore 99,6%.

5.1.2 Recognition of number of giants performed

The number of videos used in this study was 54. The algorithm described in the subsection 4.2.2 estimated the number of giants performed correctly 51 times. In three videos, there was a false positive during the kip-cast handstand at the start of the performance, leading to one giant too many recognized. The rate of correctness was therefore 51/54 $\approx$ 94%.

5.2 Validity of the body joint angle estimates

5.2.1 Body joint angle estimates based on the original frames

The results of the visual inspection of the body pose and joint angle estimates are given in the Table 5.1. Roughly 77% of frames were acceptable (good-looking or quite ok) on total level. An up-and-down estimate was present in particular at 310 degrees phase (45 cases) and at 12 degrees phase (21 cases).

Table 5.1: Results of visual inspection of the body pose and body joint angle estimates of original video frames

Phase in degrees	**12**	**40**	**70**	**110**	**145**	**270**	**310**	**Total**
Good-looking estimate	152	141	166	140	173	98	101	965
Quite OK estimate	37	66	39	11	13	69	48	283
Several angles missing from estimate	8	3	6	27	17	44	26	131
No estimate	2	1	-	18	13	7	4	45
False up and down estimate	21	15	12	13	7	1	45	114
Outsider person appearing in the frame	10	7	10	8	5	8	6	54
Something else falsely estimated as person	3	-	-	16	5	6	3	33
Total	233	233	233	233	233	233	233	1631

5.2.2 Body joint angle estimates based on the rotated and cropped frames

The results of the visual inspection of the body pose and joint angle estimates from the rotated and cropped frames are given in the Table 5.2. Roughly 92% (≈ (1425+76)/1631) of frames were acceptable (good-looking or quite ok) on total level. An up-and-down estimate was still present in particular at 310 degrees phase, but to less extent (7/233 ≈ 3%). Outsiders were cropped out from the picture as expected, and no background was estimated as human body after cropping - hence there is no column for them in the Table 5.2.

Table 5.2: Results of visual inspection of the body pose and body joint angle estimates of the rotated and cropped frames

Phase in degrees	**12**	**40**	**70**	**110**	**145**	**270**	**310**	**Total**
OK estimate	230	218	209	211	192	191	174	1425
Quite OK estimate	1	9	6	12	7	17	24	76
Several angles missing from estimate	1	6	18	10	33	22	28	118
False up and down estimate	1	-	-	-	1	3	7	12
Total	233	233	233	233	233	233	233	1631

Camera specific visual inspection results

The acceptability rates for the two video cameras used, based on visual inspection, are given both for the original and the rotated and cropped frames in the Table 5.3.

Table 5.3: Body pose and angle estimate acceptance rates for the two cameras used, based on visual inspection.

Camera frame rate and resolution	**Number of videos**	**Number of phases**	**Share of acceptable body pose and angle estimates**	
			Original frame	**Rotated and cropped frame**
60 fps, 1920 x 1080	44	1407	80%	94%
50 fps, 1280 x 720	10	227	55%	78%

Post-condition results

The angle-estimate acceptance rates, i.e. number of angle estimates that met post-conditions described in the Subsection 4.3.2 divided by all angle estimates are reported in the Table 5.4.

Table 5.4: Angle estimate acceptance rates at different phases.

Phase	**Elbow**	**Shoulder**	**Hip**	**Knee**
12 degrees	97 %	96 %	99 %	98 %
40 degrees	98 %	97 %	95 %	94 %
70 degrees	96 %	97 %	93 %	91 %
110 degrees	97 %	94 %	95 %	94 %
145 degrees	96 %	85 %	84 %	85 %
270 degrees	92 %	88 %	83 %	84 %
310 degrees	88 %	85 %	79 %	74 %
Average	94 %	90 %	87 %	86 %

Descriptive statistics of the produced and accepted angle estimates

The arithmetic averages of the body joint angle estimates per each phase and joint are illustrated in the Figure 5.1. The corresponding descriptive statistics are also given in the Table 5.5.

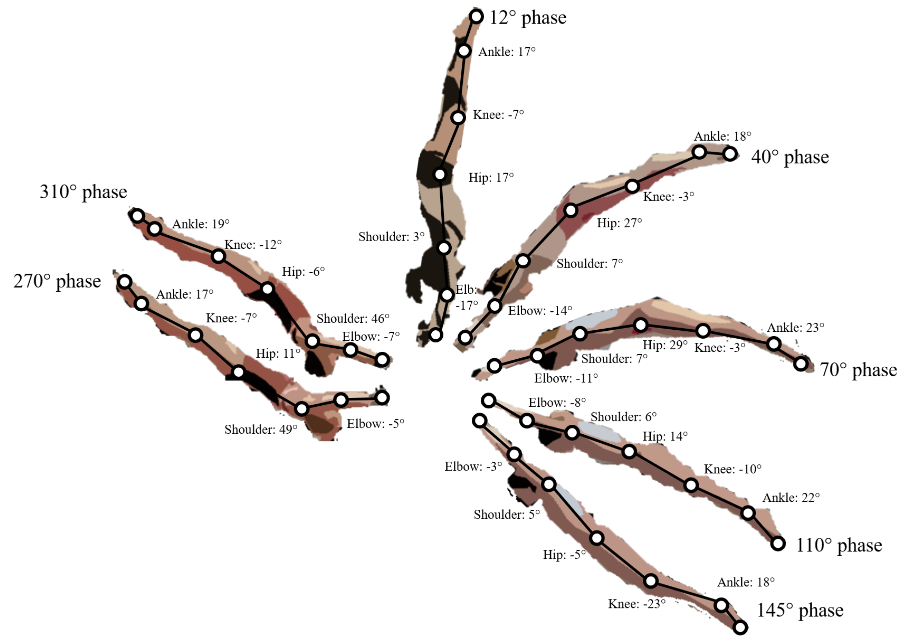


Figure 5.1: Average joint angle estimates

Table 5.5: Descriptive statistics of angle estimates

	Phase	N	M	SD	min	25% percentile	50% percentile	75% percentile	max
Elbow	12 degrees	225	−17	6.2	−28.0	−21.8	−16.3	−12.7	5.0
	40 degrees	228	−14	6.3	−27.1	−17.9	−14.3	−10.7	7.0
	70 degrees	223	−11	6.9	−29.0	−15.7	−11.5	−6.2	11.5
	110 degrees	225	−8	5.9	−21.4	−12.0	−8.3	−4.0	12.3
	145 degrees	224	−3	6.0	−20.4	−7.3	−3.1	0.3	19.2
	270 degrees	214	−5	6.4	−27.5	−8.7	−5.1	−0.7	15.5
	310 degrees	206	−7	7.0	−23.0	−11.2	−8.0	−2.9	19.3
Shoulder	12 degrees	223	3	7.1	−11.9	−0.6	2.6	7.3	47.0
	40 degrees	225	7	6.7	−8.5	3.1	6.8	11.4	34.1
	70 degrees	225	7	6.3	−11.2	2.4	8.0	11.8	21.1
	110 degrees	219	6	6.6	−11.5	0.9	6.3	10.2	25.9
	145 degrees	197	5	7.1	−10.0	0.0	4.7	9.2	23.9
	270 degrees	205	49	11.3	−14.0	44.7	49.6	54.2	68.8
	310 degrees	197	46	12.0	−7.0	42.1	46.8	52.0	78.2
Hip	12 degrees	230	17	9.7	−19.4	12.0	18.1	23.3	35.3
	40 degrees	222	27	8.0	8.0	20.4	26.3	32.9	48.6
	70 degrees	216	29	9.5	7.6	22.1	29.1	35.5	56.5
	110 degrees	221	14	11.2	−10.0	5.6	14.3	23.3	39.3
	145 degrees	196	−5	8.4	−29.2	−9.6	−5.0	1.1	16.9
	270 degrees	194	11	10.0	−16.3	5.9	11.4	18.0	33.4
	310 degrees	185	−6	9.1	−36.0	−11.0	−5.7	−0.1	18.5
Knee	12 degrees	229	−7	4.0	−17.1	−9.4	−6.4	−3.4	2.6
	40 degrees	219	−3	4.1	−16.2	−5.7	−3.4	−0.6	14.2
	70 degrees	212	−3	4.1	−20.9	−5.7	−2.8	−0.3	13.1
	110 degrees	219	−10	8.4	−54.0	−13.8	−8.8	−4.7	5.3
	145 degrees	198	−23	13.3	−73.7	−27.9	−21.4	−16.4	6.7
	270 degrees	196	−7	5.6	−23.0	−10.4	−6.8	−3.4	9.5
	310 degrees	173	−12	8.1	−67.0	−14.5	−11.3	−7.0	5.0

5.3 Angle thresholds that maximize correlation with coach evaluations

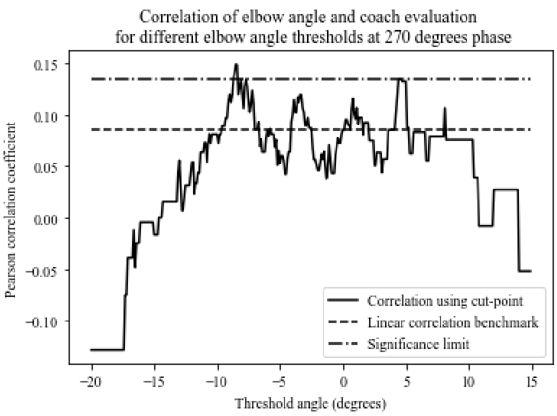
The optimal threshold angles that maximize the correlation between binary transformation of angle estimate and the correponding coach evaluation are given in the Table 5.6. The highest Pearson correlation coefficient in absolute value was $r = -0.66$ observed for knees at 145 degrees phase. All the correlation coefficients were significant, except for elbow and shoulder at 310 degrees phase. For elbows, correlation coefficient could not be calculated for 12, 40, 110 and 145 degrees, because all the performances were assessed as pure by coaches. The Figure 5.2 illustrates the Pearson correlation coefficient of binarized angles using different values of angle threshold for the four body joints at 270 degrees phase.

Table 5.6: Angle thresholds that optimize correlation with coach evaluations. r denotes Pearson correlation coefficient, N denotes number of valid observations.

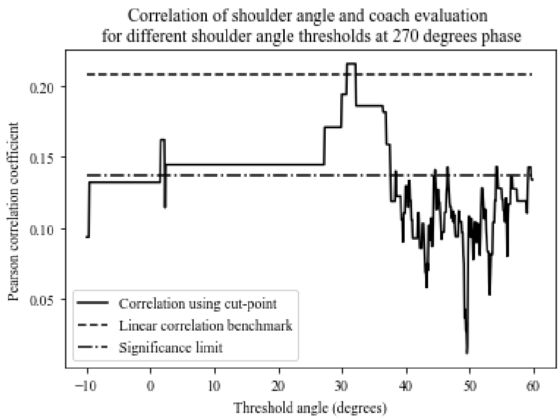
Phase	Elbow			Shoulder			Hip			Knee		
	Angle	r	N	Angle	r	N	Angle	r	N	Angle	r	N
12 degrees	N.A. (**)		225	6,6	0,21	223	0,9	-0,32	230	-6,0	-0,17	229
40 degrees	N.A. (**)		228	-1,1	-0,14	225	33,0	-0,29	222	-6,2	-0,30	219
110 degrees	N.A. (**)		225	0,2	-0,23	219	14,2	-0,26	221	-16,0	-0,37	219
145 degrees	N.A. (**)		224	8,7	-0,23	197	3,2	-0,26	196	-7,8	-0,66	198
270 degrees	-8,5	0,15	214	30,8	0,22	205	7,1	-0,27	194	-15,0	-0,52	196
310 degrees	-0,8	-0,14 (*)	206	48,1	0,14 (*)	197	-0,5	0,17	185	-24,5	-0,60	173

* Not significant

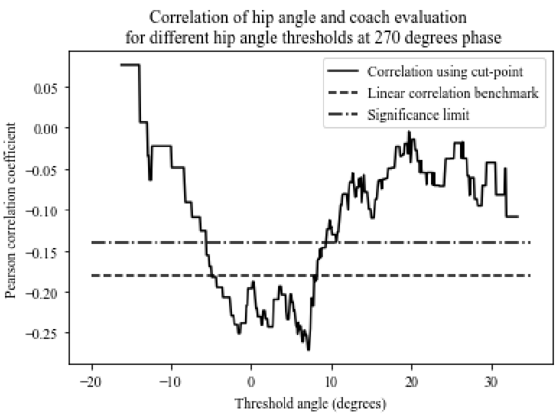
** All performances were evaluated as pure by coaches



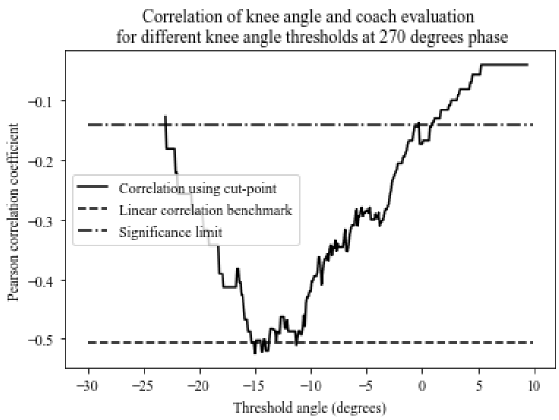
(a)



(b)



(c)



(d)

Figure 5.2: Pearson correlation coefficient of binarized knee (a), shoulder (b), hip (c) and knee (d) angle and coach evaluation for different values of threshold angle at 270 degrees phase.

6 Discussion and conclusions

The current study aimed to shed light on use of computer vision in gymnastics coaching, by exploring whether it is possible to create a system that estimates gymnasts body joint angles during a backward giant circle performance, under what conditions and to what extent the angle estimates are sensible, and how they correlate with coach evaluations.

6.1 Interpretation of the results

6.1.1 Body joint angle estimating system for artistic gymnastics

The pilot system composition summarized in the Subsection 4.2.5 proved that it is possible to create a system that based on single RGB video data estimates gymnast's body joint angles at different phases of a backward giant circle performance. However, prior knowledge about the scenery and the progression of the gymnastics skill was required. Hardest part was not to extract angles from a frame but to build a solution that detects starting and ending moment of the skill and thereafter samples the video at correct moments. The angle estimates were calculated based on pixel coordinate estimates of body joints. This was possible thanks to perpendicular positioning of video camera that captured the gymnast movement in sagittal plane.

Phase detection based on pixel grayscale value

Selecting suitable pixel and monitoring the pixel grayscale value was found to be a successful strategy to detect gymnast at a certain position, with very high 99,6% rate of correctness. The used threshold of $S = 2,4$ was an outcome of early tests, so this correctness rate should be considered more as a training measure than validation measure. There is no guarantee for generalization!

Gymnast detection based on the pixel color value did not indicate the direction of the movement. Consequently, the algorithm had to contain rules specific to the backward giant circle performance that do not generalize to other skills or other locations. The conditions were also quite favorable for the study. There was a wall behind the scene that provided with a static background. There were no windows in the room, so lightning conditions were stable most of the times, providing stable pixel value of the static background. The method had one drawback, though: the entire video file had to be scanned frame by frame from start to end to create the population of the grayscale values of the pixel, to estimate the pixel value arithmetic mean value and standard deviance. Scanning the frames of the entire file could potentially be a slowing bottleneck in a practical application where the ambition is to give feedback immediately after a repetition of the skill.

Recognition of number of giants performed

Number of backward giant circle repetitions was correctly estimated in 94% of video files, which can be considered as a successul rate. However, it is lower than the correctness rate 99,6% for the standalone phases. This stresses that it was more challenging to estimate the direction of movement than the mere position of a foreground object. The recognition algorithm required in-depth analysis of different patterns of carrying out the backward giant circle skill. Such an analysis is believed to be a pre-condition of applying artificial intelligence successfully on artistic gymnastics, as gymnastics skills differ in progression, and because gymnasts may have different ways to execute a skill.

6.1.2 Sensibility of the body joint angle estimates

Based on the results from the visual inspection of the frames and estimates, rotating and cropping of the selected frames increased the acceptability rate from 77% to 92%. Therefore, it is advisable to do the rotation and cropping in a circular movement like backward giant circle. The rotating and cropping was originally done because early tests showed that when the gymnast was at handstand, or close to handstand, OpenPose misestimated the person to be upright. Worth observing is however that rotation did not completely remove the up-and-down misestimates. There were 12 up-and-down estimates of the rotated pictures, where the gymnast was upright and ‘hanging‘. The cropping proved to be a successful method for eliminating so called false positives: an object in the background, or an outsider to be estimated instead of the gymnast. There were no such false positives after the rotation and cropping algorithm. Outsiders could have been excluded by additional on-site arrangements, but for tackling false positives from background, cropping seemed to be the only option. OpenPose had high accuracy on the original frames though, as only 2% of the frames (33 / 1631) got a complete misestimate. This is in line with observations by Nakano et al., who studied OpenPose in 3D-setting against marker-based optical motion capture and found high accuracy with some exceptions like elbow angle during a throw activity [54]. The high accuracy in the current study is believed to be partly because of no frames were chosen from the region of high speed and partial occlusion caused by supporting structures between the gymnast and the camera.

The camera properties had clear impact on the success rate of OpenPose, both on original and rotated frames, as seen in Table 5.3. The camera with higher frame rate and resolution had significantly higher probability to produce frames that get sensible body pose and angle estimates.

According to the Table 5.4, elbows had highest acceptance rate and knees lowest acceptance rate, when post-condition acceptance rules were applied. This could be because the velocity of knees was highest and the velocity of elbows was lowest. The results for rate of acceptability per phase were difficult to interpret. It seems like there are several factors influencing the quality of estimates, such as speed of motion, phase (or body rotation in wider terms), but also local lightning conditions and cloth, skin and hair color could have impact. The question, under what circumstances the angle estimates are valid, remains open. More research is required, for example by recording performances simultaneously with multiple cameras with different frame rates and resolutions, to understand the impact of video data quality on the performance of OpenPose. Only thereafter one can answer the question if there are phases of backward giant circle performance where the angle estimates are clearly different in terms of reliability. Based on the current study, one can only say that the current algorithm produces sensible angle estimates with roughly a 90% success rate for the current data set. This study aimed at exploring whether it is possible to extract angle estimates, therefore the generalization ability study of the applied methods was left out of the study.

6.1.3 Correlation between angle estimates and coach evaluation

The cross-correlations between the binary transformations of body posture angles and coach evaluations of the same body part were to a majority significant, but weak - between 0.14 and 0.66 in absolute value. The knee angle showed strongest correlation with coach evaluations during all phases. This can reflect that knee being a hinge joint, coaches truly evaluate the joint angle, whereas for hips and shoulders, coaches also evaluate the pose of lower and upper back, that are more complex than a hinge joint.

The directions of correlation made sense. For example at 45 degrees phase the coaches stressed the importance of hollow body as a strategy to prepare for passing the low bar. Consequently they saw it as fault if gymnast had hips and shoulders too extended - with low value of angle. The observed negative correlation between hip and shoulder angle and corresponding fault was in line with this.

The progress of optimal threshold of hip angle was interesting. At the handstand, the threshold was at 1 degree. Smaller angle than that was likely to indicate fault. At 40 degrees phase, angle below 33 degrees was likely to indicate fault. When passing the low bar, angle below 14 degrees was likely to indicate fault. At 145 degrees phase (arched body), one would expect that angle over a threshold would indicate fault, but reults show that angle below 3 degrees indicated fault. This can be interpreted that the arch phase would not take place yet at 145 degrees, but a bit later. After the tap phase, at 270 degrees phase, angle below 7 degrees indicates fault, underlining that hollow body is correct. Prior to final handstand, at 310 degrees phase, the correlation was only slightly over significant on 5% level, and angle over 0 degrees indicated fault, which contradicts the phase VII in the Figure 1.1 and should therefore be discarded given the low significance.

Also the shoulder angle over 31 degrees indicating fault at 270 degrees phase is worth pointing out, as it is in line with a typical performance problem in the backward giant circle pointed out by Sands in 1997 [55].

Summing up, there was no single optimal threshold value that would indicate a fault. The optimal threshold values were different for different body joints and phases. Also the directions (whether the angle should be smaller or larger than the threshold) were specific to joint and phase. This work limited the study to reporting the angle threshold values and the correlations using the entire dataset. Deeper knowledge should be gained by applying structured cross-validation methods, in order to understand how the results generalize. It should also be noted that the threshold values were specific to the group of gymnasts that participated in the study. If similar data was recorded from performances of gymnasts with different skill level, the derived angle thresholds and correlations may turn out very different.

6.2 Practical implications

For gymnastics coaches and any other reader interested in the development of the artistic gymnastics coaching, this study gives insight that it is possible to estimate body joint angles based on video feed without using markers or other sensors, mainly thanks to the development of body pose estimation algorithms and underlying research in areas of artificial intelligence, like convolutional neural networks. Data is the fuel of the artificial intelligence, therefore some practical learnings from video data gathering on-site follows.

The resolution of videos in the current study was sufficient for using computer vision algorithms. It has been earlier shown that elite gymnasts should have a technique that is robust to timing perturbations of order 20 ms [56], so 50 fps is on border being acceptable to capture flaws in timing or body pose. The experience from this study was that frame per rate of 50-60 fps was in general sufficient with exception to the areas of high speed of motion. Higher frame rate would lead to a higher quality pose recognition. Therefore, it is recommended to invest in a camera that has minimum 120 fps combined with a sufficient resolution to cover the entire orbit of the gymnast in future studies.

Overall, the study benefitted from structured arrangements at site, like steady camera positioning, markers set up on the wall behind the horizontal bars, steady lighting in the room and making sure that no other people (like coach) were in the scene while the performances were recorded. This lead to more stable conditions, and consequent possibility to tailor the computer vision algorithms using a priori knowledge. Some additional arrangements would ease the study in future: First, one or two chessboard kind of markers should be placed in the scene so that they are fully visible in every frame. This will help eliminating camera movement and also enable camera calibration. Second, there should be at least partial wider areas of background that are completely uniform in color (white). In the current study, the background was a brickwall painted in white, and a plain white paper sheet attached to the wall is believed to reduce noise when monitoring RGB-values of pixels in similar conditions, as the texture of the wall also affected the RGB-values.

The development and application of systems that truly benefit gymnasts and coaches in their daily sessions, takes place on-site at the gym. Therefore, it can be worthwhile to plan also for future technology when planning a new gym or renovating an existing one. Reserving a well-planned placement for one or several cameras can enable utilizing the latest technologies in computer vision, without disturbing the normal training routines.

6.3 Limitations

The current study used data from seven snapshots of each performance. Backward giant circle is a dynamic movement, where timing of actions carried out by the gymnast is crucial. The setup of studying only seven snapshots does not answer to the question about timing. More frequent sampling is needed to connect the use of body joint angle estimates to classical biomechanical research. To facilitate a more frequent sampling, a camera with higher frame per rate is needed.

Verification of yielded angle estimates was done through light objective visual inspection of the whole body model. In this study, scoring provided by OpenPose and manually defined postconditions were used to eliminate angle estimates before data analysis was carried out. A thorough verification, that would correspond to PCP [20], PCK or APK evaluation criterias [21], requires a time-consuming systematic annotation of every relevant body part estimate. Only if location estimates of all three body points that define an angle are seen as valid, the consequent angle is correct.

The study was based on video recorded performances of 11 gymnasts with quite equal age and skill profile. Consequently, the angle estimates that optimized the correlation with the coach evaluations reflect the age and skill profile of the participating gymnasts. The results do not necessarily generalize for gymnasts with a different skill profile. The generalizability study of the results was left to the future, because this study was limited to building understanding of the algorithms needed to produce angle estimates.

6.4 Possible future work

Next natural steps would be to continue the work to create a practical application. On that roadmap, the next step would be to design a feedback prioritization algorithm in close cooperation with gymnastics coaches. An on-site small-scale pilot could be carried out to test the practical applicability and to collect user feedback.

Gymnasts wish to focus on one correction at a time. The fact that knee angle estimates correlated strongest with coach evaluations speaks for focusing on knees first. According to gymnasts interviewed, knees are the first body part to correct, as hips and shoulders can only work correctly, if knees are fully extended. When knees are faultless, the next point of correction according to gymnasts would be the phase-specific poses: hollow, arch, tap. The current study did not try to describe the poses, but that is a strong candidate for future work. Using angle estimates, the pose could be categorized and described, in order to provide feedback that gymnasts can relate to. In the future work, it seems important to stick to analysis methods whose end results can be explained to gymnastics coaches. Therefore, building sum variables over different phases of performance, through e.g. principal component analysis, or clustering analysis of several dimensions, is not advised. Instead, building descriptive models using decision trees, or predictive analysis using binomial regression could be applicable. Data analysis could be done to find out if there are body joint angles in the early phases of backward giant circle skill that predict potential success/problem at the end of the execution of the skill. This could provide with data-driven knowledge that may help to prioritize selecting the most important feedback to the gymnast.

Other tempting future work include collecting video material with a camera that has higher frame rate, minimum 120 frames per second. Higher frame rate could help capturing gymnast's body pose in the areas of high speed. More precise frames could lead to improved body pose and joint angle estimates. Improved estimates could in turn make it possible to estimate body joint trajectories and thereby enable connecting this branch of work with classical biomechanical modeling. In additional to estimating the trajectories, higher sampling would provide knowledge about the optimal timing of gymnast's muscle activities. Combining the video analysis with simultaneous data from accelerometers attached to body and perhaps infra-red camera and reflecting markers could verify angles and increase understanding on how computer vision can be applied to estimate angular speed, momentum and forces.

Annotation of artistic gymnastics specific pictures is another field of possible future work. The objective of such a work would be to increase the body pose model performance in artistic gymnastics through training the models with augmented data.

Finally, a similar study could be carried out using a data from depth-sensing camera, something that could potentially make it possible to have the coach in the scene assisting the gymnast. That could increase the usability of the system that this work aims at contributing to.

This thesis focused on applying computer vision on a single sports performance. The fact that it is possible to set up a low-budget system that monitors and evaluates angles of human body joints during a sports performance, makes it tempting to project what other parts of the society could benefit from similar systems. Some potential fields of application could be physiotherapy and occupational health. Body pose feedback could help in rehabilitation after surgery, but it could also work as preventive technology by alarming for motion patterns or mechanical joint loadings that increase risk for osteoarthritis. The world population is ageing [57], and there is consequent increase in the need for elderly care. Affordable posture feedback systems could at best contribute to increased welfare both within sports and in the entire society.

References

[1] Fédération Internationale de Gymnastique. (2020). “Presentation of Women’s Artistic Gymnastics”, [Online]. Available: www.gymnastics.sport/site/pages/disciplines/pres-wag.php (visited on 05/13/2020).

[2] M. L. Manning, “Biomechanics of technique selection in women’s artistic gymnastics”, Ph.D. dissertation, Cardiff Metropolitan University, Oct. 2014.

[3] Fédération Internationale de Gymnastique. (Dec. 2016). “2017 - 2020 Code of Points Women’s Artistic Gymnastics”, [Online]. Available: www.gymnastics.sport/publicdir/rules/files/en_WAG%20CoP%202017-2020.pdf (visited on 05/13/2020).

[4] P. Laitinen, “Backward Giant Circle Data Set: Documentation of the Procedures For Collecting an Artistic Gymnastics Element Data Set For Computer Vision and Machine Learning Purpose”, TUCS Technical Reports, no. 1207, 2020.

[5] M. R. Yeadon and M. J. Hiley, “Optimum technique for generating angular momentum in accelerated backward giant circles prior to a dismount”, *Journal of Applied Biomechanics*, vol. 19, pp. 119–130, 2003. DOI: 10.1123/jab.19.2.119.

[6] A. Simmons and S. Chappell, “Artificial intelligence-definition and practice”, *IEEE journal of oceanic engineering*, vol. 13, no. 2, pp. 14–42, 1988, ISSN: 0364-9059.

[7] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[8] E. Davies, "CHAPTER 1 - Vision, the Challenge", in *Machine Vision (Third Edition)*, ser. Signal Processing and its Applications, E. Davies, Ed., Third Edition, Burlington: Morgan Kaufmann, 2005, pp. 1–15, ISBN: 978-0-12-206093-9. DOI: https://doi.org/10.1016/B978-012206093-9/50004-6.

[9] W. Gong, X. Zhang, J. Gonzàlez, A. Sobral, T. Bouwmans, C. Tu, and E.-h. Zahzah, "Human Pose Estimation from Monocular Images: A Comprehensive Survey", *Sensors (Basel, Switzerland)*, vol. 16, no. 12, pp. 1–39, 2016, ISSN: 1424–8220. DOI: 10.3390/s16121966.

[10] G. Hidalgo, Z. Cao, T. Simon, S.-E. Wei, Y. Raaj, H. Joo, and Y. Sheikh. (Nov. 2020). "OpenPose v1.7.0", [Online]. Available: github.com/CMU-Perceptual-Computing-Lab/openpose (visited on 01/01/2021).

[11] G. Hidalgo, Y. Raaj1, H. Idrees, D. Xiang, H. Joo, T. Simon, and Y. Sheikh, "Single-Network Whole-Body Pose Estimation", *ArXiv e-prints*, 2019. arXiv: 1909.13423.

[12] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3D human pose annotations", in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1365–1372. DOI: 10.1109/ICCV.2009.5459303.

[13] L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting People Using Mutually Consistent Poselet Activations", in *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 168–181, ISBN: 978-3-642-15567-3.

[14] L. Pishchulin, M. Andriluka, P. Gehler, and B. Schiele, "Strong Appearance and Expressive Spatial Models for Human Pose Estimation", in *2013 IEEE Interna-*

tional Conference on Computer Vision, 2013, pp. 3487–3494. DOI: 10.1109/ICCV.2013.433.

[15] A. Hernández-Vela, S. Sclaroff, and S. Escalera, "Poselet-Based Contextual Rescoring for Human Pose Estimation via Pictorial Structures", *International Journal of Computer Vision*, vol. 118, no. 1, pp. 49–64, 2016, ISSN: 0920-5691. DOI: 10.1007/s11263-015-0869-y.

[16] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014.

[17] M. P. I. Informatik. (2014). "MPII Human Pose Dataset", [Online]. Available: human-pose.mpi-inf.mpg.de (visited on 02/02/2021).

[18] S. Johnson and M. Everingham, "Learning Effective Human Pose Estimation from Inaccurate Annotation", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[19] S. Johnson. (2011). "Leeds Sports Pose", [Online]. Available: sam.johnson.io/research/lspet.html (visited on 02/02/2021).

[20] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation", in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587468.

[21] Y. Yang and D. Ramanan, "Articulated Human Detection with Flexible Mixtures of Parts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2878–2890, 2013. DOI: 10.1109/TPAMI.2012.261.

[22] A. Newell, K. Yang, and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation", in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 483–499, ISBN: 978-3-319-46484-8.

[23] G. Hua, L. Li, and S. Liu, "Multipath affinage stacked-hourglass networks for human pose estimation", *Front. Comput. Sci. 14*, Apr. 2020. DOI: 10.1007/s11704-019-8266-2.

[24] E. P. Ijjina and K. M. Chalavadi, "Human action recognition in RGB-D videos using motion sequence information and deep learning", *Pattern Recognition*, vol. 72, pp. 504–516, Jul. 2017. DOI: 10.1016/j.patcog.2017.07.013.

[25] E. E. Cust, A. J. Sweeting, K. Ball, and S. Robertson, "Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance", *Journal of Sports Sciences*, vol. 37, no. 5, pp. 568–600, 2019. DOI: 10.1080/02640414.2018.1521769.

[26] M. P. Díaz-Pereira, G.-C. Iván, M. Escalona, and D. N. Olivieri, "Automatic recognition and scoring of olympic rhythmic gymnastic movements", *Human Movement Science*, vol. 34, pp. 63–80, 2014. DOI: 10.1016/j.humov.2014.01.001.

[27] M. Mack, M. Bryan, and T. Heinen, "Modeling judges' scores in artistic gymnastics", *The Open Sports Sciences Journal*, vol. 12, pp. 1–9, 2019. DOI: 10.2174/1875399X01912010001.

[28] P. Nevalainen, A. Kauhanen, C. Raduly-Baka, M.-J. Laakso, and J. Heikkonen., "Video based Swimming Analysis for Fast Feedback", in *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM*, INSTICC, SciTePress, 2016, pp. 457–466. DOI: 10.5220/0005753704570466.

[29] M. Einfalt, D. Zecha, and R. Lienhart, "Activity-Conditioned Continuous Human Pose Estimation for Performance Analysis of Athletes Using the Example of Swimming", in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 446–455. DOI: 10.1109/WACV.2018.00055.

[30] A. Nibali, Z. He, S. Morgan, and D. Greenwood, "Extraction and Classification of Diving Clips From Continuous Video Footage", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jul. 2017, pp. 38–48.

[31] R. Kanth, J.-P. Skön, K. Lehtomäki, P. Nevalainen, P. Virjonen, M.-J. Laakso, and J. Heikkonen, "Image Analysis and Development of Graphical User Interface for Pole Vault Action", *Journal of Image and Graphics*, vol. 6, no. 2, pp. 109–116, 2018. DOI: `10.18178/joig.6.2.109-116`.

[32] H. Fujiwara and K. Ito, "ICT-based Judging Support System for Artistic Gymnastics and Intended New World Created Through 3D Sensing Technology", *Fujitsu Scientific & Technical Journal*, vol. 54, no. 4, pp. 66–72, Sep. 2018. [Online]. Available: `www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol54-4/paper11.pdf` (visited on 04/29/2021).

[33] The International Gymnastics Federation (FIG), Fujitsu Limited. (Oct. 2019). "'A step towards the future' with the first official use of Fujitsu technology to support judging at the 2019 Artistic Gymnastics World Championships", [Online]. Available: `www.fujitsu.com/global/about/resources/news/press-releases/2019/1002-01.html` (visited on 04/19/2021).

[34] F. Potdevin, O. Vors, A. Huchez, M. Lamour, K. Davids, and C. Schnitzler, "How can video feedback be used in physical education to support novice learning in gymnastics? Effects on motor learning, self-assessment and motivation", *Physical education and sport pedagogy*, vol. 23, no. 6, pp. 559–574, 2018, ISSN: 1742-5786. DOI: `10.1080/17408989.2018.1485138`.

[35] G. Irwin, S. Hanton, and D. G. Kerwin, "The conceptual process of skill progression development in artistic gymnastics", *Journal of Sports Sciences*, vol. 23, no. 10, pp. 1089–1099, 2005. DOI: 10.1080/02640410500130763.

[36] M. L. Manning, G. Irwin, M. Gittoes, and D. Kerwin, "Influence of longswing technique on the kinematics and key release parameters of the straddle Tkatchev on uneven bars", *Sports Biomechanics*, vol. 10, no. 3, pp. 161–173, 2011.

[37] W. Witten, "A kinematic and kinetic analysis of the overgrip giant swing on the uneven parallell bars", Ph.D. dissertation, Michigan State University, Apr. 1990.

[38] M. R. Yeadon and M. J. Hiley, "The mechanics of the backward giant circle on the high bar", *Human Movement Science*, vol. 19, pp. 153–173, 2000. DOI: 10.1016/S0167-9457(00)00008-7.

[39] M. J. Hiley and M. R. Yeadon, "Optimization of high bar circling technique for consistent performance of a triple piked somersault dismount", *Journal of Biomechanics*, vol. 41, pp. 1730–1735, 2008. DOI: 10.1016/j.jbiomech.2008.02.028.

[40] M. J. Hiley and M. R. Yeadon, "Achieving consistent performance in a complex whole body movement: The Tkatchev on high bar", *Human Movement Science*, vol. 31, pp. 834–843, 2012. DOI: 10.1016/j.humov.2011.08.009.

[41] G. Williams, G. Irwin, D. Kerwin, and K. Newell, "Changes in joint kinetics during learning the longswing on high bar", *Journal of sports sciences*, vol. 33, pp. 1–10, May 2014. DOI: 10.1080/02640414.2014.921831.

[42] A. L. Sheets and M. Hubbard, "Evaluation of a subject-specific female gymnast model and simulation of an uneven parallell bar swing", *Journal of Biomechanics*, vol. 41, pp. 3139–3144, 2008. DOI: 10.1016/j.jbiomech.2008.08.027.

[43] M. Hiley and F. Yeadon, "The margin for error when releasing the asymmetric bars for dismounts", *Journal of Applied Biomechanics*, vol. 21, pp. 223–235, 2005. DOI: 10.1123/jab.21.3.223.

[44] Canon. (2018). "Canon EOS 1300D Specifications", [Online]. Available: www.canon.co.uk/for_home/product_finder/cameras/digital_slr/eos_1300d/specification.html (visited on 05/05/2020).

[45] DJI OSMO. (Sep. 2019). "Osmo Pocket User Manual v1.2", [Online]. Available: dl.djicdn.com/downloads/Osmo%20Pocket/20190919/Osmo_Pocket_User_Manual_v1.2_en.pdf (visited on 05/05/2020).

[46] M. W. Berry, A. Mohamed, and B. W. Yap, *Supervised and Unsupervised Learning for Data Science*, 1st. Springer Publishing Company, Incorporated, 2019, ISBN: 3030224740. DOI: 10.1007/978-3-030-22475-2.

[47] Fédération Internationale de Gymnastique. (Jan. 2017). "FIG General Judges' Rules, Version 1.2 January 2017", [Online]. Available: https://www.gymnastics.sport/publicdir/rules/files/en_WAG%20Specific%20Judges%20Rules%202017-2020%20.pdf (visited on 05/13/2020).

[48] International Telecommunication Union. (Mar. 2011). "Recommendation ITU-R BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios", [Online]. Available: www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf (visited on 04/28/2021).

[49] M. Piccardi, "Background subtraction techniques: a review", in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 4, 2004, 3099–3104 vol.4. DOI: 10.1109/ICSMC.2004.1400815.

[50] D. D. Bloisi, “Background Modeling and Foreground Detection for Maritime Video Surveillance”, in *Handbook on Background Modeling and Foreground Detection for Video Surveillance: Traditional and Recent Approaches, Implementations, Benchmarking and Evaluation*, T. Bouwmans, F. Porikli, B. Hörferlin, and A. Vacavant, Eds., Chapman and Hall/CRC, 2014, pp. 14-1–14-22, ISBN: 978-1-4822-0537-4. DOI: 10.1201/b17223-18. [Online]. Available: www.dis.uniroma1.it/~bloisi/papers/background-modeling-and-foreground-detection-for-maritime-video-surveillance.pdf.

[51] J. Han, M. Kamber, and J. Pei, “12 - Outlier Detection”, in *Data Mining (Third Edition)*, ser. The Morgan Kaufmann Series in Data Management Systems, J. Han, M. Kamber, and J. Pei, Eds., Third Edition, Boston: Morgan Kaufmann, 2012, pp. 543–584, ISBN: 978-0-12-381479-1. DOI: doi.org/10.1016/B978-0-12-381479-1.00012-5.

[52] P. Laitinen. (2020). “Gymnastics Analytics”, [Online]. Available: gitlab.utu.fi/gymnasticsanalytics (visited on 02/28/2021).

[53] B. Meyer, “Applying "Design by Contract"”, *Computer*, vol. 25, no. 10, pp. 40–51, Oct. 1992, ISSN: 0018-9162. DOI: 10.1109/2.161279.

[54] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, “Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras”, *Frontiers in Sports and Active Living*, vol. 2, p. 50, 2020, ISSN: 2624-9367. DOI: 10.3389/fspor.2020.00050.

[55] W. A. Sands, “A Typical Performance Problem in a Giant Swing”, *Technique*, vol. 17, no. 1, pp. 1–2, 1997. [Online]. Available: usagym.org/pages/home/publications/technique/1997/1/giant.pdf (visited on 05/02/2021).

[56] M. J. Hiley and M. R. Yeadon, "Optimisation of backward giant circle technique on the asymmetric bars", *Journal of Applied Biomechanics*, vol. 23, pp. 300–308, 2007. DOI: 10.1123/jab.23.4.300.

[57] UN, Population division, "World population ageing, 2019 : highlights", 37 p. : 2019. [Online]. Available: http://digitallibrary.un.org/record/3846855 (visited on 05/05/2021).

Appendix A Example pictures

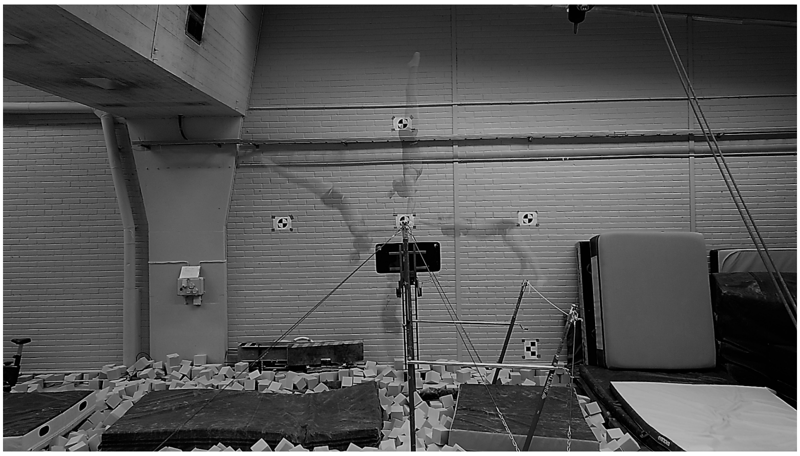


Figure A.1: Video scene example with four snapshots from one performance blended on top of each other.

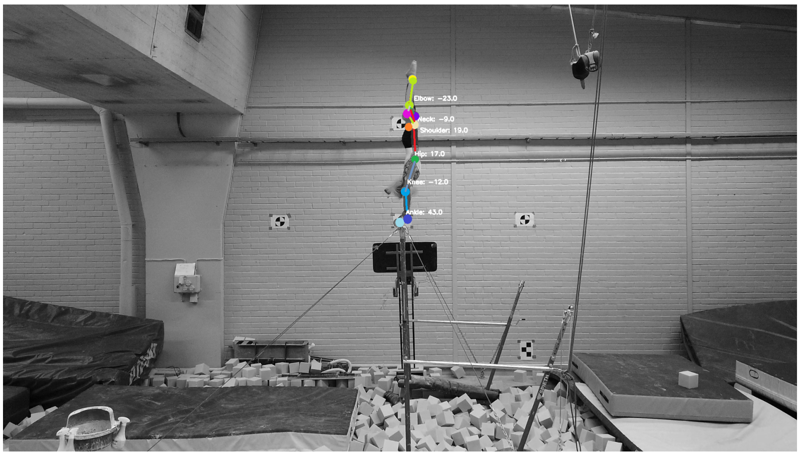


Figure A.2: Example frame where Openpose algorithm has detected a person but misestimated the joint positions. The frames were rotated to avoid such misestimates.

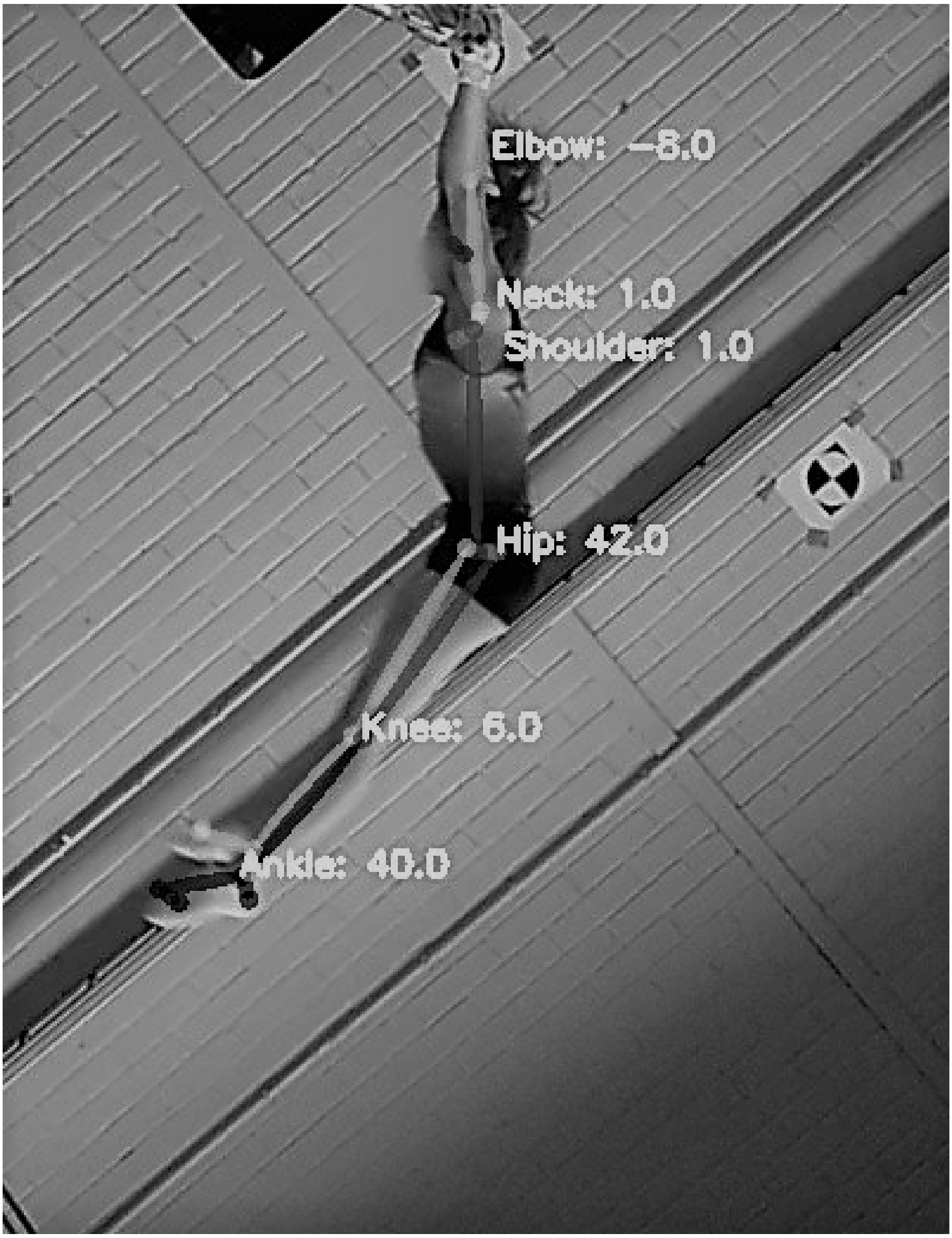


Figure A.3: Example of a rotated and cropped frame, where the gymnast is at 40 degrees phase.

Appendix B OpenPose body25 model and angle definitions

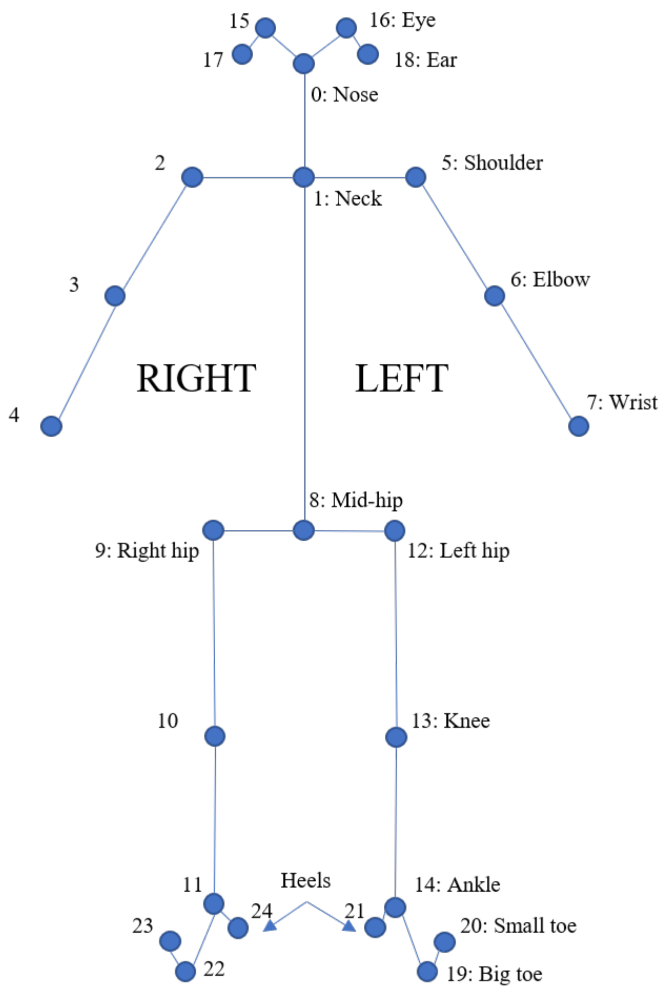


Figure B.1: Keypoint numbering of OpenPose's body25 model option - from a frontal view.

Table B.1: Body points used when calculating the joint angles.

Angle	Joint A		Joint B		Joint C	
	OpenPose number	Description	OpenPose number	Description	OpenPose number	Description
Elbow	7	Left wrist	6	Left elbow	5	Left shoulder
Shoulder	6	Left elbow	5	Left shoulder	12	Left hip
Neck	18	Left ear	1	Neck	12	Left hip
Hip	1	Neck	12	Left hip	13	Left knee
Knee	12	Left hip	13	Left knee	14	Left ankle
Ankle	13	Left knee	14	Left ankle	19	Left bigtoe
Leg Split*	22	Right bigtoe	12	Left hip	19	Left bigtoe
Body angle within frame **	(7)	Left wrist offset vertically	7	Left wrist	12	Left hip

* Legs fully together was defined as 0 degrees

** Body angle was calculated using left wrist offset vertically (y-coordinate forced to 0) as point A