MEASUREMENT OF ELITE TAEKWONDO ATHLETES USING CONVOLUTIONAL NEURAL NETWORKS IN COMPETITION ENVIRONMENTS

Shaun Douglas Barratt

A thesis submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Doctor of Philosophy

December 2025

Candidate Declaration

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.

2. None of the material contained in the thesis has been used in any other submission for an academic award.

3. I certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged. I confirm that I have sought and obtained copyright permission for any third party materials included in this thesis. I confirm that no Generative AI tools were used in the preparation or completion of this assessment. This submission aligns with AITS 1 of the Artificial Intelligence Transparency Scale (AITS).

4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy, and ethics approval has been granted for all research studies in the thesis, as shown in the table below.

Ethics review reference number	Title of research study	Approval date
ER39637393	Player event detection within combat sport using neural networks and computer vision	02/02/2022

5. The word count of this thesis is 32,000.

Name	Shaun Douglas Barratt
Date	December 2025
Award	PhD
College	Health, Wellbeing and Life Sciences
Director of Studies	Dr Simon R Goodwill

Abstract

Elite Taekwondo performance analysis requires valid and scalable ways to measure athlete positions and interpersonal distances directly from competition video. This research study explores whether convolutional neural networks (CNNs) can achieve such measurements in authentic, single-camera environments that include occlusions, referees, and changing viewpoints. To answer this, first this study examined what manual annotation can achieve: expert annotators provided the benchmark for reliability, and elite coaches classified distance categories to assess their qualitative agreement and how these related to real-world ranges. Agreement between coaches was only moderate, showing broad overlap in perceived distance bands, suggesting that fixed thresholds are unsuitable without coach-specific calibration.

Building from this manual baseline, the extent to which automated systems could replicate or improve upon these capabilities was analysed. Manual annotation demonstrated high reliability in favourable viewing conditions but deteriorated with greater occlusion and shallower camera angles. These findings defined the standard against which automation was assessed.

A custom YOLO detector and subsequent pose estimation pipeline were developed to automate athlete localisation and distance estimation. Detection provided a robust basis for identifying athletes, while pose estimation refined positional accuracy by capturing body landmarks. Combining the two within a hybrid framework reduced identity swaps and stabilised distance estimates across diverse scenarios. Overhead or elevated audience views supported the most consistent results, whereas broadcast footage with heavy occlusion remained challenging.

Overall, 2D CNN-based measurement in elite Taekwondo proved it was able to accurately measure when supported by careful calibration and a scenario-aware processing pipeline. Automated systems can objectively extract metrics with the equivalent accuracy of manual reliability under optimal viewpoints, though precise distance inference remains difficult in complex visual conditions. Practically, these tools should report continuous distance metrics and, where categorical outputs are desired, apply coach-calibrated mappings to ensure interpretive validity. This approach aligns automated performance analysis with human perception while accommodating the visual and tactical complexities of real competition.

Acknowledgements

I would like to express my deepest thanks to my Director of Studies, Dr Simon R Goodwill, for his support, guidance, encouragement, and insight throughout this thesis. I would also like to thank my supervisors Dr Sergio Davies, Dr Andy Hext and Dr Chaung-Yuan Chui for their excellence in their respective fields and input throughout the PhD.

I would like to extend these thanks to all staff and researchers within the Sports Engineering Research Group at Sheffield Hallam University, and those at the Advanced Wellbeing Research Centre, in particular Katie Mills, Daniel Haid, Ben Heller, Carole Harris, Lisa Jones, Rea Smith, Christy Bannister each of you contributed so much to this PhD and helped to progress its future.

Thank you to all the individuals, athletes, analysts and coaches who kindly volunteered to participate in the studies conducted throughout this programme of research, especially Garry Owen and Mario Contarino, providing me with large amounts of high-level rich data that was vital to the success of this project, your support will always be greatly appreciated.

Finally, I would like to thank my family and friends for their continued encouragement and support, Dad & Charles. A special thanks to Nicola Hatton, who during this thesis, became Nicola Barratt, for spending countless hours reading and re-reading this thesis and its studies, even if most of it “sounds like gibberish”, as well as being an emotional rock for my first international conference. And to our son, George Barratt (now four): thank you for your 3 a.m. peer-review comments, especially for their brevity, volume, and excellent use of caps.

Contents

Abstract i

Acknowledgements iv

Contents v

Nomenclature vii

Chapter 1 – Introduction 1

Chapter 2 – Literature Review 7

2.1 Introduction 7
2.2 Taekwondo 8
2.3 Manual versus Automated Methods for Movement Measurement 13
2.4 Modern CNN-Based Approaches to Athlete Tracking 18
2.5 Camera Calibration and Distance Measurement Techniques 24
2.6 Challenges in Real-World Taekwondo Vision Analysis 31
2.7 Summary of gaps and rationale 33

Chapter 3 – Data collection 35

3.1 Introduction 35
3.2 Video Footage 35
3.3 Athlete Position Measurements 44
3.4 Conclusion 51

Chapter 4 – Understanding the domain 52

4.1 Introduction 52
4.2 Methodology 53
4.3 Results 58
4.4 Established Framework 70
4.5 Discussion 71
4.6 Conclusion 73

Chapter 5 – Reliability of Manual Measurement 75

5.1 Introduction 75
5.2 Method 76
5.3 Results 79
5.4 Discussion 83
5.5 Conclusion 86

Chapter 6 – Coach Distance Reliability **87**

6.1 Introduction 87
6.2 Method 87
6.3 Results 90
6.4 Discussion 97
6.5 Limitations 97
6.6 Conclusion 98

Chapter 7 – Bounding Box Automation **99**

7.1 Introduction 99
7.2 Methodology 100
7.3 Results 104
7.4 Conclusion 116

Chapter 8 – Pose Estimation Automation **118**

8.1 Introduction 118
8.2 Methodology 119
8.3 Results 122
8.4 Discussion 126
8.5 Conclusion 128

Chapter 9 – Discussion **129**

9.1 Manual Annotation Reliability and Manual Benchmarking 129
9.2 Coach Distance Classification 130
9.3 CNN-Based Bounding Box Tracking Performance 131
9.4 Pose Estimation Model Performance 133
9.5 Combining Bounding Boxes and Pose Estimation 134
9.6 Measurement Challenges and Contextual Factors 136
9.7 Implications for Applied Sports Science Analysis 138
9.8 Future Research 141
9.9 Conclusions 143

References 143

Nomenclature

Abbreviation	Meaning (as used in this research study)
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CoM	Centre of Mass
COCO	Common Objects in Context
CVAT	Computer Vision Annotation Tool
DLT	Direct Linear Transformation
fps	Frames Per Second
GPS	Global Positioning System
HOG	Histogram of Oriented Gradients.
HOTA	Higher-Order Tracking Accuracy
ICC	Intraclass Correlation Coefficient
IMU	Inertial Measurement Unit
IoU	Intersection over Union
JSON	JavaScript Object Notation
LPS	Local Positioning System
mAP	Mean Average Precision
MMA	Mixed Martial Arts
MOTA	Multi-Object Tracking Accuracy
MS COCO	Microsoft Common Objects in Context
PSS	Protective Scoring System
RCNN	Region-based Convolutional Neural Network
SD	Standard Deviation
SSD	Single Shot Detector
UV	Image Coordinate System
XYZ	World Coordinate System
WT	World Taekwondo
YOLO	You Only Look Once

Chapter 1 – Introduction

Elite sports demand precise measurement of athlete movements to inform training and performance enhancement. In disciplines ranging from team sports to combat sports, the ability to quantify positioning, distance and motion of athletes provides valuable feedback for coaches and athletes [1]. Combat sports such as Taekwondo hinge on split-second movements and tactical spacing. Measuring how athletes move relative to each other can reveal insights into strategy effectiveness, injury risk, work-rate and an athlete's ability to execute tactical actions. Sports scientists increasingly recognise that objective motion-based data, such as distances covered, timing of attacks and positioning, are essential for understanding performance and tailoring coaching interventions. Capturing such data in an elite competition setting is inherently challenging. Traditional motion capture methods, such as optical marker-based tracking systems are impractical in live competitions, and manual video analysis is labour intensive and prone to inconsistencies [1]. This motivates the exploration into modern computer vision techniques, especially Convolutional Neural Networks (CNNs), to automatically track athlete movements from standard video footage.

Within sports science and performance analysis, Taekwondo has the unique demands in that it requires accurate movement measurement due to its close-combat setting. Taekwondo is a fast-paced Olympic combat sport where athletes constantly adjust their distance to execute kicks and avoid opponent attacks. Prior research in Taekwondo performance analysis has highlighted that the interpersonal distance between athletes is a critical metric. For example, Maloney *et al*. noted that distance is a universally useful indicator of engagement across all kick types. In their study, distance was defined as the "tracking of the centre of mass, i.e. the midpoint between fighters' feet" [2].

The aim of this research study is to validate whether convolutional neural networks (CNNs) can be used to measure athlete movements during Taekwondo competition.

Previous studies in other sports have employed pose estimation and bounding box models (Figure 1.1), assessing their precision and recall, but these have rarely been applied to real-world competition footage in Taekwondo nor evaluated by the various scenarios within the sport.

Figure 1.1 - Bounding box detections in sports - FootyVision multi-object tracking [3]

Within Taekwondo specifically, pose estimation has been tested on athletes in competition areas [4]. This has largely been in controlled laboratory environments as illustrated in Figure 1.2, In such settings, athletes are often not in correct protective uniform, and conditions do not account for real-world factors such as false detections, referee interference, background clutter, or the effects of clothing and protective garments as shown in Figure 1.3. OpenPose in particular is prone to assigning keypoints incorrectly when multiple individuals are present, limiting its reliability [5].

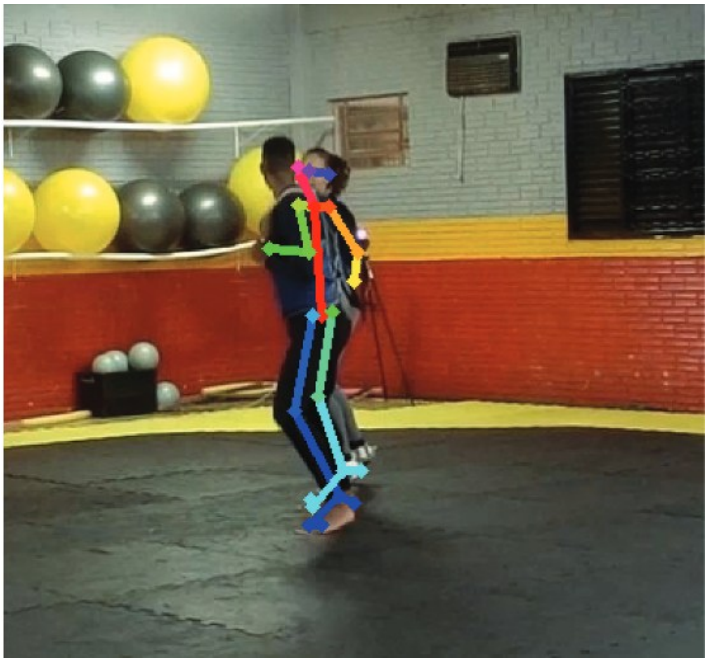


Figure 1.2 - OpenPose errors in detecting multiple people due to occlusion [5]

Figure 1.3 – OpenPose model applied in a competition environment

Crucially, there is limited research investigating the use of CNNs for extracting positional data from historical Taekwondo competition and training footage, despite this footage being widely available and previously used to inform performance analysis studies in Taekwondo [2].

This research study was motivated by the needs of GB Taekwondo, the national governing body responsible for developing elite athletes who compete at Olympic and World Championship level. Performance analysts and coaches within GB Taekwondo require objective, scalable methods to quantify athlete movement and tactical positioning during competition. Traditionally, analysis has relied on subjective observation or labour-intensive manual video annotation, both of which limit the volume of data that can be processed and introduce potential inconsistencies across analysts. GB Taekwondo identified interpersonal distance measurement as a priority metric, as it directly informs tactical decision-making, training interventions, and post-competition feedback. However, existing performance analysis workflows lacked a validated, automated system capable of extracting this metric from the extensive archive of competition footage already captured at major tournaments.

The potential value of such a system extends beyond retrospective analysis. If proven valid and reliable, automated measurement could enable GB Taekwondo coaches to rapidly evaluate athlete performance across multiple bouts, compare tactical approaches between weight categories, and identify patterns associated with successful outcomes. Additionally, automated tracking would support longitudinal monitoring of athlete development, providing quantitative evidence of tactical maturation or regression over training cycles. By validating convolutional neural network (CNN) based methods against manual benchmarks and coach-defined standards, this research study aimed to establish whether such technology could meet GB Taekwondo's operational requirements for accuracy, robustness, and practical deployment in real competition environments. The findings therefore have direct implications for how the national programme structures its performance analysis infrastructure and integrates objective movement data into coaching practice.

To address this gap, the research study first benchmarks manual performance through manual annotation and coach-defined classifications, before systematically testing bounding box and post estimation approaches. This staged approach ensures than CNN-based systems are not only tested in technical isolation but also evaluated against the reliability of manual annotation and the practical knowledge of coaches.

Objectives

To achieve the overall aim, the research is guided by the following objectives:

1. **Dataset Preparation**

   Compile a dataset of historical Taekwondo competition footage for analysis (Chapter 3)

2. **Scenario Framework**

   Develop a data-driven framework from the dataset preparation to categorise annotation scenarios such as viewpoint, orientation and occlusion to enable a stratified error analysis based in the sport (Chapter 4).

3. **Manual reliability benchmark**

   Assess the intra- and inter- operator reliability of manual annotation methods, providing a baseline for automated systems (Chapter 5).

4. **Coach Classifications**

   Evaluate the consistency of coach-defined distance categories and establish thresholds for automated measurement systems (Chapter 6).

5. **Bounding box models**

   Assess CNN-based bounding box models for containment and positional accuracy, including the development of a custom model using transfer learning for athlete identification (Chapter 7).

6. **Pose estimation models**

   Assess CNN-based pose estimation models and evaluate a hybrid approach that combines bounding box identification with pose data for improved reliability (Chapter 8).

7. **Applied integration and implications**

   Integrate findings from manual and automated analyses to develop practical recommendations for performance analysis workflows, highlighting the real-world implications for applied Taekwondo research and coaching practice (Chapter 9).

Chapter 2 – Literature Review

2.1 Introduction

In this literature review, relevant works and methods were examined to situate this research within the domains of sports performance analysis and computer vision-based motion tracking. The review begins with an overview of Taekwondo as a sport and the performance metrics that analysts and coaches focus on, establishing why movement measurement is important.

Methods used to track and measure athletes' positions were surveyed, from traditional manual annotation techniques to IMU sensors and vision-based approaches. This research is motivated by the need to understand interpersonal distance in historical competition footage. Accordingly, the literature review focuses on vision-based methods, examining both traditional computer vision techniques and deep learning models for object detection and pose estimation in sport. Recent key studies, including Taekwondo as well as other sports such as Judo, Boxing, Football are examined and discussed to understand the limitations in these systems.

Challenges inherent in real-world competition footage, such as occlusion, motion blur and other visual clutter are highlighted as these factors influence the design and expected performance of any automated system.

Finally, the gap in current literature is identified, highlighting the lack of real-world validated CNN-based measurement in real Taekwondo competition environments.

Although this chapter provides an overview of existing methods for athlete tracking, it also identifies several methodological and contextual weaknesses in the literature. Many studies validate computer vision models under ideal laboratory conditions, with limited

testing in authentic competition footage. Consequently, while prior work demonstrates technical feasibility, its practical reliability in complex, real-world sports environments remain uncertain.

2.2 Taekwondo

Taekwondo as a sport today has become recognised globally, both through World Taekwondo (WT) and the Summer Olympic Games. Taekwondo is the national martial art of South Korea, while the origin of Taekwondo has been debated for many years as it has been heavily influenced by the Japanese during their occupation of Korea in the early 1900's [6], as well as having connections to the China's wushu. Today it is believed that Taekwondo is a fusion of several different sources, or it is a form of martial art that places emphasis on foot skills, distinguishing it from Japan's Karate and China's Wushu [7]. In a survey conducted by the Korean government, taekwondo is practiced by 70 million people in over 190 countries worldwide [8].

Taekwondo first began its move towards global status after the 1950s through the efforts of Korean immigrants and martial artists. A large part of its success in the United States of America (USA) was primarily due to Jhoon Rhee, who has been coined 'father of American Taekwondo' for opening the first Taekwondo school in America [7], [9], [10], [11].

World Taekwondo (WT) - formerly the World Taekwondo Federation is an organisation founded in 1973 to promote the sport internationally [12], WT became responsible for the forms used, training issues, competition rules, credentials, and promotions of the sport globally. WT was successful in lobbying to include the sparring element of Taekwondo training as an official sport, first recognised at the 2000 Sydney Olympic Games [12].

Since then, the professionalism and competitiveness has substantially increased, particularly at the higher levels.

Although this outlines the historical and cultural development of Taekwondo, it also highlights an important gap in the academic literature. The globalisation of the sport and its transformation into an Olympic discipline have not yet been matched by a body of empirical research on performance measurement or movement analysis. Much of the available literature on Taekwondo's history is descriptive or sociocultural, with limited data-driven or biomechanical studies. As such, while the evolution of Taekwondo is well documented, there remains a clear need for systematic, quantitative research into the sport's tactical performance aspects. A gap this research study aims to address through the application and validation of modern data-capture methods.

2.2.1 Point Scoring

Within Taekwondo the focus is to technical knockout or score more points than the opponent through various punches to the body and kicks to the head or body. To gain a point certain criterion has to be met as per the World Taekwondo Competition Rules [13].

The WT rules highlight several key considerations, firstly that there are only two valid areas of contact for the athlete to score, this means that athletes must focus on precision and power in their strikes. Secondly, the technique used and how it is delivered makes a difference to the points scored, for example a turning kick to the head is worth 5 points whereas a punch to the trunk is worth a single point. Additionally, a gam-jeom is to be avoided as it gives the opponent a point. At the end of the match the athlete with the most points win. A gam-jeom may be given for several reasons, such as exiting the contest area or falling, these are discussed further in the sub-chapter "Prohibited acts". Finally, this criterion also leads a suggestion to tailor the performance towards each specific

weight category, gender and age grouping as the WT Technical Committee tailor the sensitivity of the Protective Scoring System (PSS) to these factors.

Taekwondo contains many specialist types of kicks and as such have been studied in detail for example, the kick forces based on the male roundhouse kick [14] as well as the effect of distance within roundhouse kicks [15], [16], however not all kicks have had the opportunity for this in-depth analysis.

WT also specify several prohibited acts within sparring Taekwondo each one carries a penalty of 1 point also referred to as a gam-jeom. If an athlete intentionally and repeatedly refuses to comply with the competition rules, then the referee can end the match by raising a yellow card and declaring the opposing contestant the winner.

If an athlete receives 10 gam-jeom's the referee declares the athlete, the loser by punitive declaration (PUN).

2.2.2 Context and Performance Metrics

Taekwondo bouts take place on a fixed octagonal or rectangular mat of 8 metres by 8 metres, where two athletes attempt to score points by landing kicks (and punches, to a lesser extent) on the opponents the head, or torso known as the scoring zones. As kicking is the predominant attack, and as kicks can only score with sufficient force, keeping an appropriate distance between an opponent, known as distance management, is a fundamental aspect of Taekwondo strategy [11 - 15]. Accordingly, coaches and analysts pay attention not only to the number and type of techniques, but also to how athletes position themselves and move relative to the opponent over the course of a match.

Previous research has investigated Taekwondo actions and their tactical context. Menescardi *et al*. [19] developed a Taekwondo combat model using Markov analysis to represent sequences of attack and counter-attack, highlighting common patterns in high-

level competitions. This kind of modelling shows that it is not just individual techniques that matter, but the spatio-temporal structure of exchanges, which can be unique to each set of athletes, this is also known as the action – reaction system. Similarly, work by Maloney *et al*. examined the role of distance in engagement, by using a simulated combat setup to qualify how distance impacts technique effectiveness [2]. In this study, distance was formally defined by tracking the athlete's centre of mass (midpoint between the feet) as illustrated in Figure 1.1. This has then been used in increasing studies as the accepted definition to measure distance in Taekwondo [5]. This builds upon work performed by Headrick *et al*'s work of measuring distance in football [22]. Maloney's findings reinforced that regardless of the specific kick executed, the distance between athletes is a critical metric for understanding the dynamics of the two athletes.

Figure 2.1 – Estimated CoM position

Other studies have measured specific Taekwondo techniques (e.g. roundhouse kick impacts) in relation to distance, the definition here is slightly different "The target distance was defined as the horizontal distance between the big toe of the front (left) foot and centre of the target." [18]. This shows that quantifying movement and positioning

can yield insights into performance (such as optimal range for power or scoring). Additionally other studies into effects of actions in Taekwondo have measured from the pelvis [23], or the centre of the target area [15].

Athlete position is frequently used in sports literature as a key component for measuring athlete distance [23 - 25]. Within Taekwondo, it has been applied to establish external training load metrics, assess tactical ability, and evaluate interpersonal dynamics [8] - [10]. While some studies employ continuous measurements through digital analysis of specific points, others rely on qualitative descriptors such as clinch, short, medium, and at length. Although these terms are generally consistent in relation to leg lengths [15], [20], the actual distances they represent can vary depending on coaching interpretations. An example of these distance classifications is provided in Figure 3.11.

Figure 3.11 – Coach distance classifications in Taekwondo

In summary, the sports science literature establishes why measuring movement in Taekwondo is important: it is closely tied to technical and tactical outcomes. What is less established is how to measure those movements efficiently in real match environments. Traditional approaches in Taekwondo and similar sports have relied on either manual observation or laboratory setups, each with limitations. The subsequent sections review these approaches, from manual video analysis techniques to sensor-based systems and, importantly, to the computer vision methods (particularly CNN-based) that are the focus of recent developments.

2.3 Manual versus Automated Methods for Movement Measurement

2.3.1 Manual Annotation methods in Sports

Before modern machine learning techniques, performance analysts often measured athlete movements through manual video annotation, sometimes referred to as coding, notational analysis, or motion analysis [28], [29], [30], [31]. In practice this involved stepping through video frames and marking athlete positions by hand or using software tools to digitise coordinates on the screen. For example, an analyst might pause footage at regular intervals and record the location of each athlete on the mat to estimate distances, or trace movement paths. Common tools such as CVAT [32], [33], Kinovea [33], [34] and Dartfish [34] have been employed to assist with manual digitisation tasks. These applications allow a video to be calibrated to real-world dimensions and then used to measure the distance by drawing lines or tracking points frame-by-frame [33]. In Taekwondo, an analyst could calibrate the 8x8 metre competition area and then manually trace the athletes' positions over time, obtaining their trajectories and interpersonal distance.

Manual video analysis can be accurate when performed carefully. Studies in field sports report that experienced operators digitising video can achieve high reliability [35], [36]. In one rugby analysis, intra-operator variance was reported at under 0.5% and inter-operator variance under 1% for total distance measured [1]. This suggests that given a clear protocol and calibration, independent analysts can produce consistent results for distances and speeds.

Manual measurements are also time-consuming, subjective and prone to human error. Maintaining consistency over long periods of time is difficult, and biases can arise from individual judgement, such as when exactly to mark a position, or how to handle occluded

frames. Research across combat sports has shown that rater fatigue and ambiguous visual frames can increase error rates over time [37], [38]. For more complex annotation tasks, such as classifying movement types or assessing qualitative distances, studies have found only moderate agreement between observers, which error margins often ranging between 5% and 10% depending on the metric [39]. In Taekwondo an additional challenge is the speed of movements, a single kick can be just a fraction of a second, which requires the use of frame-by-frame tracking which can be arduous and inconsistent when motion blur occurs, which is common in broadcast footage.

Despite these limitations, manual annotation has served as the de facto standard in many sports analysis studies, including Taekwondo. It has the advantage of human judgement, as an analyst can recognise which athlete is which, discount irrelevant individuals in the scene, and interpret events in context. Tasks that still challenge automated systems.

Manual annotation remains constrained by its dependence on consistency and labour intensity. Reported reliability figures often come from controlled environments and may not translate to noisy, multi-actor competition footage. Few studies quantify how operator fatigue or visual ambiguity affects measurement accuracy, leaving uncertainty about its robustness in long-duration analyses. Therefore, while manual methods provide essential benchmarks, they are not scalable for large datasets and are inherently limited in their repeatability. Automated systems must therefore demonstrate comparable accuracy while improving efficiency and standardisation if they are to be considered viable replacements.

2.3.2 Sensor-based and other methods

Aside from video analysis, other measurement approaches have been explored in sports, though their applicability to Taekwondo is limited, and in competition, even more so. In open-field sports such as football and rugby, teams commonly use Global Positioning Systems (GPS) or Local Positioning Systems (LPS) worn by athletes to track running

distance and speed. These systems can provide real-time data over large fields [40], but they require athletes to wear a device (usually a small unit in a vest) and are accurate for larger translational movement, often around 0.5m, they are not appropriate for the centimetre level accuracy required from that of combat sports, especially in an indoor setting, where GPS signals can be subject to interference.

Another alternative is to use inertial measurement units (IMUs) or other wearable motion sensors. Previous studies have used IMUs on limbs to capture kick velocities or identify techniques [41], [42], [43], [44], additionally electronic impact sensors are already built into Taekwondo protective gear for use in scoring. Some coaches believe that this impacts the effectiveness of the protective gear to provide sufficient safety. Additionally, these sensors do not directly measure the distance between two athletes, nor their global position on the competition area. IMUs generally focus on specific actions or impacts. Multi-sensor approaches (combining IMUs, with other sensors in the arena such as force plates) have been proposed, but such setups are complex, not currently available and are not typically deployed in Taekwondo [45].

The gold-standard for motion capture is marker-based optical tracking [46] (e.g. Vicon, Qualisys & OptiTrack), these systems can produce extremely accurate 3D kinematics by placing retro reflective markers on the athlete's body and using multiple high-speed cameras. While marker-based systems are used in lab biomechanical studies such as measuring a Taekwondo kick under controlled conditions [47], they are not viable in live competitions. Athletes can not wear markers or suits in bouts due to safety, and competitions take place in arenas sometimes with uncontrolled lighting and backgrounds. Marker less systems like the one explored in this research study aim to replicate some of the capabilities of motion capture without those constraints.

In summary, non-vision methods either lack the granularity needed (GPS gives a broad position, but not precise enough for two athletes in close quarters) or lack practicality for competition (IMUs and markers interfere with natural competition, or only capture part of the picture). This reinforces the use of computer vision as a promising solution as video is readily available for most events as Taekwondo matches are routinely filed for officiating and broadcasting. Modern algorithms in computer vision potentially can extract the needed data from these existing videos, without the need of specialist equipment placed on athletes.

2.3.3 Traditional Computer Vision Techniques

Before machine learning algorithms became dominant, traditional computer vision methods were applied to track athletes in sports footage with some success. Many earlier systems relied on background subtraction and object segmentation to detect athletes [48]. The idea is to separate moving persons (foreground) from the static or slowly changing background in scenarios like a fixed camera on a playing field or court. Additionally, techniques such as frame differencing or colour histogram thresholding can then be applied to identify blobs corresponding to athletes [49]. In football video analysis, research often used colour-based segmentation, since the field is green and athletes wear distinctive kits, it is possible to isolate athlete blobs by filtering out the green background. Those blobs might then be tracked over time with algorithms like Kalman filters, particle filters or mean-shift tracking [50]. In this study athletes were represented as collections of particles, which could help maintain an estimated position even through brief occlusions. Multi-camera setups are also used to mitigate occlusion issues, as a athlete out of view of one camera may be detected by another, systems then often integrate these views to keep track.

Applying these traditional methods to historical Taekwondo footage presents difficulties. Unlike a field sport with a uniform backdrop, Taekwondo competitions have a more variable background, (referees moving, multi-coloured mats, logos, audience in view). The athletes' uniforms are often similar in colour as both athletes wear white doboks, with only the chest protectors differing in red vs blue, making colour segmentation less straightforward, simple background subtraction could struggle when referee or others enter the camera frame, potentially misidentifying them as athletes. Additionally, fast movements such as spinning kicks can lead to motion blur, causing shape distortions that basic blob detectors might not handle well.

Another traditional technique is using feature descriptors like Histogram of Oriented Gradients (HOG) combined with a traditional classifier like SVM. For instance, Baysal *et al*. used HOG+SVM to detect and track athletes in football footage [51]. A similar detector could in theory be trained for Taekwondo athletes, but without deep learning, its accuracy would likely be limited by the variety of poses and orientations the athletes can assume. Overall, while traditional computer vision methods laid important groundwork, they tend to require extensive tuning and still have difficulties when applied to the complex visuals of bouts in Taekwondo, rapid motion, occlusion between athletes, and interference from other visual artifacts remain challenging in this field.

In summary, traditional computer vision approaches can detect and track athletes under certain conditions but they have clear limitations in robustness. They struggle particularly with Occlusions, Overlap and Object differentiation. These limitations set the stage for modern deep learning approaches, which have started to overcome many of these challenges through learned representations and data-driven models.

2.4 Modern CNN-Based Approaches to Athlete Tracking

With the introduction of deep learning, and convolutional neural networks (CNNs) in particular, computer vision tasks like object detection, tracking and pose estimation in the last decade has been revolutionised. CNNs can automatically learn rich feature representations from images, enabling for more reliable detection of objects, such as athletes, under a diverse range of conditions than fined-tuned traditional computer vision methods could achieve [52], [53], [54], [55]. In this section, two key CNN-driven techniques relevant to measuring athlete movement are reviewed: object detection (bounding box tracking) and pose estimation. How these are evaluated, using metrics such as mAP, and how they are applied in sports contexts are detailed.

2.4.1 CNN-Based Object Detection and Tracking

Object detection CNNs take an image as an input and output bounding boxes around objects of interest, usually along with class labels. In terms of this research study, the object of interest are people, specifically athletes. State-of-the-art detectors can find people in images even with complex backgrounds, this is due to the models often being pre-trained on large datasets such as MS COCO [56] a sample of this dataset is shown in Figure 2.2. These datasets contain thousands of people in varied poses and environments.

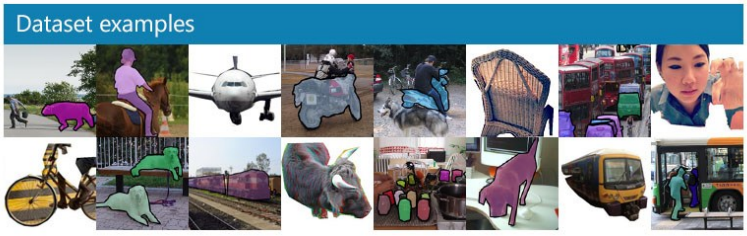


Figure 2.2 - MS COCO Examples

Two broad families of detectors are widely used: one-stage detectors and two-stage detectors.

- **One-stage detectors** such as You Only Look Once (YOLO) [53], [57], [58], Single Shot Detector (SSD) [59], RetinaNet [60] predict bounding boxes and class probabilities in a single network pass, prioritising speed. For example, the latest YOLO models are known for real-time performance, YOLOv7, used in the FootyVision system [3], can process high resolution football frames while maintaining a high accuracy.
- **Two-stage detectors** such as Faster RCNN [61], Mask R-CNN [62] and Cascade R-CNN [63] first generate region proposals, then refine classifications in a second stage. They often achieve higher accuracy at the cost of speed. Faster R-CNN remains a popular choice when detection precision is paramount. In a research context with offline analysis of footage, a two-stage detector to maximise detection of athletes, especially in challenging frames might be appropriate. Two-stage models are more computationally intensive, which can be a factor given the number of frames in a dataset.

Selecting an appropriate detector depends on the specific task requirements. In this research study, as high accuracy in distinguishing two athletes and a referee at minimum, a customised approach was taken (Chapter 7). By using custom training, it is possible to boost performance for domain-specific cases while still using a One-stage detector. For example, FootyVision trained a YOLOv7 model on football-specific data (players and ball) and achieved a high mean average precision (mAP 95.7%) and excellent tracking metrics (MOTA ~94%) on football videos. Similarly in combat sports, Quinn and Corcoran trained a YOLOv5 model to detect boxers and MMA fighters in competition

footage, achieving around 95% mAP for boxing match detection [64]. These results suggest that with sufficient training data and proper network selection, CNN can surpass earlier methods, handling fast motions and cluttered backgrounds.

One challenge in object detection for sports is maintaining identity tracking across frames, i.e. knowing that the athlete in the red protecting equipment at time t is the same as the one in red at time $t + 1$. Detection alone processes each frame independently; therefore, to obtain trajectories, a multi-object tracking (MOT) algorithm is often used. Common MOT approaches (SORT [65], DeepSORT [66], ByteTrack[67], BotSORT [68], etc.) link detections frame to frame based on motion consistency and appearance features. In combat sports, where only two primary athletes are present, MOT is on the face much simpler than tracking 22 football players. When trialled, these tracking systems often fail if the two subjects swap places or come into contact, which happens frequently in the case of Taekwondo. The custom detection approach in this work (labelling each athlete separately) is effectively a built-in data association, the network itself distinguishes “blue” from “red” athlete, eliminating the need for a post MOT system in many cases. This technique of using distinct class labels or uniform colours to main identity has precedent, in football, it is possible to train separate detectors for home vs away teams if kits are consistent, in Quinn’s MMA work, the authors explicitly used red / blue to identify athlete identities.

When evaluating models for detection and tracking there are many metrics to choose from. The mean Average Precision (mAP) is a de facto standard metric for object detection, summarising precision-recall curves for the detector across classes. High mAP (close to 1.0 or 100%) means the detector finds most objects with few false positives. As discussed, specialised models in sports can achieve mAP in the mid-90s percent range for athlete detection [3], [64]. For tracking metrics such as Multiple Object Tracking

Accuracy (MOTA) and Higher Order Tracking Accuracy (HOTA) are used to quantify how well continuous tracks are formed. While this research study is not focused on pushing the boundary of multi-object tracking algorithms themselves, these metrics are used to ensure that the bounding box based approach is performing adequately on the general domain. The automated tracking of athletes yields accurate trajectories that can feed into distance calculations without losing track of an athlete or confusing the two athletes during occlusion.

In summary, CNN detectors provide a powerful tool for automated measurement: they can locate athletes in each frame robustly. By tailoring these models to the Taekwondo domain using transfer learning on annotated competition images, and by using strategies to preserve identity, such as identifying athletes based on their protector colour, obtaining a foundation for measuring movement, essentially a time series of co-ordinates for each athlete on the competition area. This approach reduces each person to a point or region which can then be cross compared with an analyst's annotation to be compared of accuracy.

2.4.2 Pose Estimation in Sports

Pose estimation involves detecting key skeletal points (joints like: ankles, knees, hips, shoulders, etc) on the human body from images or video. In sports contexts, pose estimation can yield detailed biomechanical data without markers, enabling analysis of technique and form. Several CNN-based frameworks for 2D pose estimation have emerged, notably OpenPose [68 - 71], AlphaPose [73], Hourglass networks [74] and HRNet [75], among others. These models are trained on datasets with keypoint annotations and can predict the pixel coordinates of around 18-25 (varies based on architecture) body landmarks for any person in an image as shown in Figure 2.3.

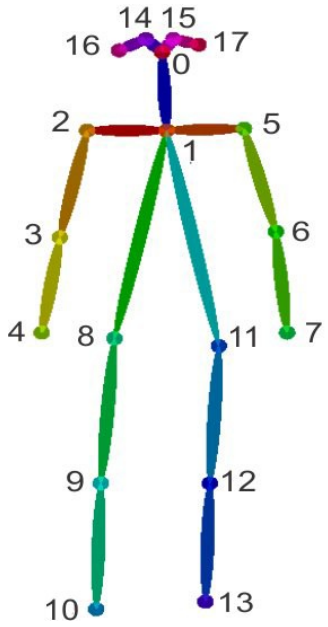


Figure 2.3 - Keypoints of a pose model used in OpenPose [69]

For Taekwondo and other martial arts, pose estimation has been explored in scenarios like training assistance and form evaluation. Poomsae (the technical forms in Taekwondo) have been analysed using pose data to judge movement correctness [76]. In sparring, a few studies have attempted to use pose estimation to track athletes, one very recent study by Banks *et al*. assessed the validity of OpenPose for tracking Taekwondo athletes position on the mat [4]. In this controlled experiment, athletes performed movements with and without an opponent, while being recorded by a 7-camera motion capture system for ground truth. The results were encouraging: the marker less OpenPose system achieved RMS errors of 10 - 30cm for 2D position of the centre of mass compared to the gold-standard, and errors of 5 – 16cm for 3D reconstructed points. Additionally, the agreement was excellent (ICC > 0.90), indicating that pose estimation can indeed be a viable tool for tracking positions or inter-athlete distances under the right conditions. Those conditions, importantly, included a high-quality video setup (two 120 Hz cameras) and a clear view

of the participants without protective gear, on an official mat, with no crowd or extraneous people and without the requirement to identify athletes.

Applying pose estimation to real competition footage introduces additional difficulties. Unlike a staged experiment, competition video may have suboptimal camera angles (not every exchange is perfectly viewed), variable frame rates and resolutions, and interference from other people such as referees, judges, coaches and spectators. OpenPose and similar multi-person models will detect all people in the frame, so if a referee is visible, the algorithm might output three sets of skeleton keypoints, two athlete and one referee, without knowing which is which. This is problematic when trying to measure the interpersonal distance between the two athletes as without identity tracking the referee may be mistaken for an athlete. Previous attempts to use pose data in a multi-person setting have had to make simplifying assumptions, such as that only two persons are present or manually filtering outputs. Some researchers have combined pose estimation with tracking or identification techniques. For example, in an MMA analysis context, one could use pose estimation to recognise certain moves (punches, kicks) while using an object detector to keep track of which athlete is which [64], similar to the hybrid method proposed in earlier sections.

Another challenge is occlusion and entanglement: when athletes are in a clinch or during a throw, their limbs may overlap, causing pose models to confuse joint locations between the two athletes or to lose some points. Advanced pose estimators like HRNet have improved robustness to occlusion by using stronger backbone networks [77], but errors can still occur. Additionally, Taekwondo athletes wear protective gear: headguards, protectors, which can sometimes obscure joints. These sport-specific attire issues can introduce small biases in keypoint detection.

Despite these challenges, pose estimation remains extremely valuable because it provides fine-grained data beyond what a simple bounding box can. For instance, using pose keypoints it is possible to measure the exact distance between athletes lead foot, or the angle of a kick, information that a bounding box cannot offer. Pose data also enables the calculation of each athlete's centre of mass (CoM) if needed, by using a weighted combination of joint positions and body segment masses. Some literature in sports has focused on CoM tracking. In Taekwondo Maloney's definition of distance was essentially the distance between athletes CoM (and further simplified, to the midpoint between athletes' feet) [78]. A pose model could directly provide an estimate of CoM for each athlete by calculating these joint positions.

Accuracy versus practicality is a key consideration with pose estimation. In ideal lab settings, it can rival manual or even marker-based accuracy [4]. In noisy real settings, some accuracy may be lost, however this field is advancing new models and algorithms (including those leveraging Transformers and 3D pose estimation) are being actively researched [79]. It is likely that robust pose tracking of athletes will continue to improve. For the purpose of this research study, the approach is to use pose estimation, but to combine it with the reliability of object detection to avoid identity swaps or false positives. This combined approach is part of the novel contribution of the work in adapting CNN tools to the Taekwondo context

2.5 Camera Calibration and Distance Measurement Techniques

Regardless of whether manual, bounding boxes or pose keypoints is used to locate athletes in video frames, a crucial step is to convert those image co-ordinates into real-world measurements. This requires understanding the camera's model and the geometry of the competition area. Most sports analysis that derives physical metrics from video involve some form of camera calibration, or homography computation. In multi-camera

systems such as Hawk-Eye, a full 3D calibration is done so that the athletes' co-ordinates can be mapped onto a model of the field or court [80]. For a single-camera approach (as is common with broadcast or historical footage), typically a planar homography is used to map positions on the 2D video frame to positions on the 2D ground plane, the competition area.

In Taekwondo competitions, the competition area's dimensions are standardised, and its shape is either square or an octagon of known size, with the competition area being 8 by 8 metres as shown in Figure 2.4.

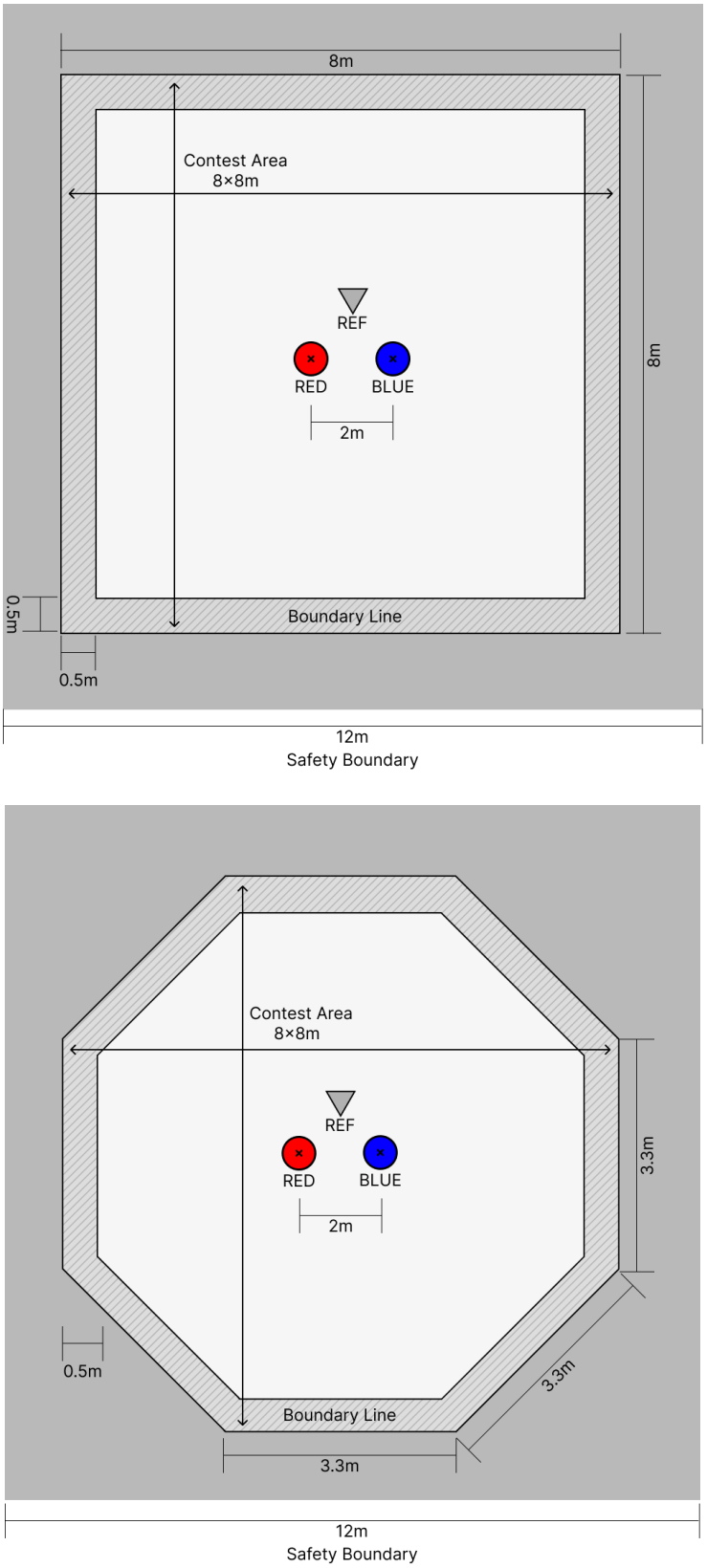


Figure 2.4 - Taekwondo Competition Area

These known dimensions can provide reference points for calibration using Direct Linear Transformation (DLT). The DLT method transforms measured line correspondences into a homogeneous system of linear equations, where the coefficients are arranged into a measurement matrix [81]. The DLT method is particularly suited to this research due to its ability to handle perspective distortions caused by varying camera angles and positions. The technique relies on a set of reference points, both in the image and in the real world, to compute a transformation matrix that maps between the two coordinate systems, the output of such method is illustrated in Figure 2.5.

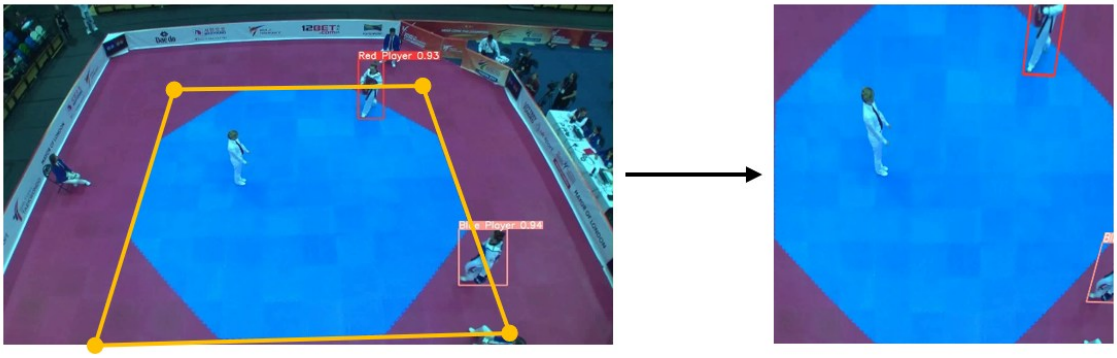


Figure 2.5 - DLT applied in Taekwondo using 4 keypoints

The accuracy of the DLT method is highly dependent on the quality and distribution of the reference points. Ideally, the reference points should be spread across the entire competition area to minimise errors caused by perspective distortions. Additionally, as a Taekwondo competition area can have 8 vertices when using the octagonal shape, these can be used to provide a higher accuracy calibration as illustrated in Figure 2.6 using A1 through A8 clockwise. By identifying these points in the video footage, a transformation can be applied to map pixel coordinates to real-world coordinates.

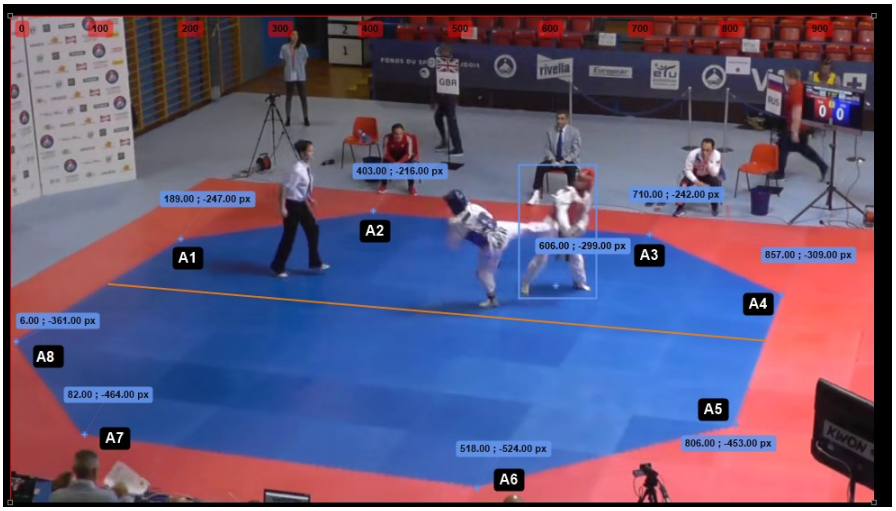


Figure 2.6 - Using 8 vertices of the competition area

In cases where the camera angle is high (e.g. a semi-bird's-eye view), a planar homography works well for mapping positions on the competition area. If the camera angle is extremely low such as in broadcast footage or athletes move in the air (jump kicks), there can be some vertical parallax error – but for measuring inter-athlete distances on the mat plane. An assumption is made that athletes remain in contact with the ground plane most of the time for distance measurement purposes.

In summary, camera calibration bridges the gap between image analysis and meaningful sport metrics. Without using camera calibration, it is only possible to get the distances in pixels, which would not consistent between bouts. By using the known geometry of the Taekwondo competition area, this approach grounds all measurements in real-world units, making the results directly interpretable for coaches and researchers. This calibration step is a vital component of validating the measurement.

2.5.1 Expected Calibration Accuracy in Taekwondo Applications

The accuracy of DLT-based position reconstruction for planar calibration is influenced by several factors, including the number and spatial distribution of calibration points, camera resolution, lens distortion, and the angle of the camera relative to the competition area. The original formulation of the DLT by Abdel-Aziz and Karara [82] established the mathematical foundation for transforming image coordinates to real-world positions using known reference points. In contemporary sports applications, homography-based calibration (a 2D planar variant of DLT) has become standard practice for mapping athlete positions from broadcast or fixed-camera footage onto a known playing surface.

Recent research in sports field calibration provides benchmarks for expected accuracy. In football, Theiner *et al*. [83] developed the TVCalib method, which minimises segment reprojection error to achieve camera calibration for broadcast video. Their evaluation on the SoccerNet dataset demonstrated that modern calibration methods can achieve reprojection errors below 5 pixels at 960×540 resolution (quarter HD), translating to approximately 10–15 centimetres in real-world coordinates on a standard football pitch. Similarly, Magera *et al*. [84] established a universal protocol for benchmarking camera calibration in sports, noting that systems achieving sub-10 pixel reprojection error at HD resolution are considered suitable for performance analysis applications. These findings suggest that for an 8-metre by 8-metre Taekwondo competition area, calibration errors in the range of 5–15 centimetres are achievable under optimal conditions when using distributed reference points and broadcast or elevated viewpoints.

Camera perspective significantly affects calibration accuracy and subsequent positional measurements. Overhead or elevated viewpoints, where the camera axis approaches perpendicular alignment with the mat plane, minimise perspective distortion and produce more uniform error distributions across the calibrated area. Broadcast or audience-level

viewpoints, characterised by shallow camera angles, introduce greater depth-axis errors due to perspective foreshortening. Empirical studies in padel and football have quantified this effect: Javadiha et al. [85] reported that high-angle cameras positioned 7.6 metres above the court yielded standard deviations below 5 cm in lateral positioning (X-axis) and 12 cm in depth positioning (Y-axis), whereas broadcast angles with lower elevation increased depth errors by a factor of two to three. This pattern reflects the inherent challenge of depth reconstruction from monocular views, where the viewing angle directly determines the sensitivity of depth estimates to pixel-level measurement noise.

The choice and distribution of calibration points also influence reconstruction accuracy. Using the four corners of the competition area provides a minimal calibration set, sufficient for basic homography estimation, but this configuration can produce increased interpolation errors towards the centre of the mat where most athlete interactions occur. By contrast, using all eight vertices of the octagonal Taekwondo competition area improves spatial coverage and reduces reconstruction error in the central region.

Given the competition footage analysed in this research study, which includes broadcast, audience, and overhead viewpoints, calibration accuracy is expected to vary according to these established patterns. Overhead footage with eight-point calibration should achieve positional errors in the range of 5-10 centimetres, comparable to accuracies reported in related sports applications. Broadcast footage with four-point calibration is anticipated to exhibit larger errors, particularly along the depth axis, in the range of 15-30 centimetres, consistent with findings from football and padel studies.

2.6 Challenges in Real-World Taekwondo Vision Analysis

Having reviewed the methodologies, it is important to detail the specific challenges that arise when applying these methods to real competition footage of Taekwondo. Many of these challenges have been briefly discussed in the current chapter. Here these challenges are consolidated and highlighted as the key difficulties the current research must investigate.

2.6.1 Occlusions and Fast Motion

Taekwondo is by nature a fast-paced sport, kicks can exceed speeds that cause motion blur on standard cameras (especially older footage with lower shutter speeds). This blur can in turn reduce detection and pose accuracy. Additionally, the two athletes frequently occlude each other, either partially or fully from the cameras view during the bout. This can cause automated detectors to cause the athletes to become undetected or for the pose estimation to lose keypoints. This is also a problem for manual annotations, where the occluded athlete may be harder to annotate when occluded.

2.6.2 Visual Complexity and Clothing

Unlike in a lab setting, competition videos have referees stepping in and out of the athletes, additionally, there are coaches, judges, spectators and sometimes cameras flashing. False-positive detections are commonplace and existing setups often mistake referees for athletes. The athletes' attire also is taken into consideration. Both athletes wear white uniforms with similar head and torso protectors, with only the protective clothing being red or blue as the main distinguishing feature. Current research often operates under lab environments where athletes are wearing plain clothes. As both athletes are wearing white, it is possible for pose detectors to mis-identify arms and legs especially when athletes perform kicking actions where their body is extended. In older

footage where the quality is much lower, the background contrast to the athletes may also be sub-optimal.

2.6.3 Multiple Persons and Identity Switching

As previously discussed, current multi-object detection algorithms simply identify persons in a frame, it is then up to a MOT algorithm to track these athletes, this still requires a manual annotation to say which athlete is which at the beginning. MOT algorithms are also not suitable for occluded close combat scenarios, as often athletes in proximity (during a clinch) may spin around, as one becomes occluded, most current tracking systems fail to track both athletes and the identities are switched. Additionally, but to a lesser extent, as most algorithms simply detect persons, false detections can arise if a judge, or the audience comes into frame, either by standing up or walking across, a basic model would tag them as a person of interest.

2.6.4 Noise and Video Quality

Historical footage can vary in quality, some videos may be interlaced or lower frame rate, causing jitter or additional preprocessing. Noise can also come in the form of compression artifacts, common when videos are downloaded from streaming platforms. Such factors can degrade the performance of CNN models, which are typically trained on clean images.

2.6.5 Generalisation across scenarios

Taekwondo competitions can take place in different venues with different lighting, mat colours, camera positions etc. A model trained on one set of videos might not immediately generalise to another, if the camera in a new video is placed at a different angle or if the mat is a different colour that affects contrast. Ensuring that the system works across

various competition scenarios, national tournaments, international opens, Olympic games etc, is a challenge and a requirement.

2.6.6 Summary of challenges

By recognising these challenges, it is possible to understand the design choices in literature and in this research. For instance, in the FootyVision study on football, the difficulties of motion blur with athletes and similar uniforms causing occlusions and identification issues [3]. This is analogous to the use case in this research study, in the original study the researchers addressed this by using a combination of customised models, and integration of several techniques. Similarly, the MMA tracking study noted that limited data and the need to classify more object types, such as gloves and shorts for identity could improve their system, indicating that domain-specific cues are important in training models for combat sports [64].

2.7 Summary of gaps and rationale

The reviewed literature shows that automated tracking has been performed in laboratory contexts, but validation on authentic Taekwondo competition footage remains limited. Methods based on bounding boxes and pose have rarely been evaluated on broadcast or audience video that contains referees, crowd movement and variable occlusion. Integrated pipelines that combine detection, identity maintenance and pose to handle two-athlete interactions are largely untested. Direct comparisons with manual annotation are rare, and error is typically reported as a single aggregate rather than stratified by sport-specific scenarios. Coaching constructs for interpersonal distance are widely used, yet there is no agreed numerical mapping between categories and real-world distances. These gaps motivate the present research, which validates convolutional neural network-based measurement on competition footage, compares manual and automated methods within a

shared scenario framework, and reports outputs that are interpretable by coaches. This rationale aligns with the research studies aim to determine whether convolutional neural networks (CNNs) can be used for positional measurement in elite Taekwondo. As such this research study addresses the following research questions:

1. Under competition conditions, what is the reliability of manual positional annotation, in terms of intra- and inter-operator agreement, when stratified by viewpoint, athlete alignment and containment as defined by the scenario framework?
2. What is the consistency with which coaches apply distance categories, and what real-world distance ranges correspond to each category when mapped to calibrated positional data?
3. What positional accuracy is achieved by a bounding box approach across viewpoints, athlete alignment and containment, and under which conditions the method is viable for applied analysis?
4. Does pose estimation, used alone or in combination with bounding box identity assignment, improve positional accuracy relative to the bounding box only approach within the same scenarios?

Chapter 3 – Data collection

3.1 Introduction

This chapter describes the construction of the dataset used throughout this research. It details the selection of historical Taekwondo competition footage, the inclusion and exclusion criteria applied, and the procedures used for manual annotation and digitisation.

The dataset was designed to capture a representative range of competition scenarios, including variation in viewpoint, resolution, athlete positioning, and occlusion. These data form the foundation for subsequent analyses of manual annotation reliability (Chapters 5–6) and the validation of automated tracking methods (Chapters 7–8).

3.2 Video Footage

The video footage used in this research was provided through collaboration with GB Taekwondo. Access was granted to an extensive archive of 38,529 videos, which encompassed recordings from numerous competitions (n = 431) and training sessions (n = 2,329), spanning the period from 2014-2024. These videos featured footage of 10,688 unique athletes across various service levels, competition types and weight categories.

The archive was comprised of footage collected from various sources, predominantly collected by GB Taekwondo and World Taekwondo, with a small portion from online sources within the public domain. This resulted in a wide range of video cameras being used, as such there are many variations in camera angle, camera position, framerate, athletes, resolutions and overall quality. All procedures were approved by Sheffield Hallam University Research Ethics Board (ER39637393).

3.2.1 Viewpoints

As illustrated in Figure 3.1, the historic footage included recordings from a variety of camera locations and angles. This variation is critical because it reflects real-world conditions under which athlete movement data is captured, ensuring the dataset's applicability to both research and practical analysis. Within competitive settings, certain scenarios occur more frequently and are more repetitive than others. Camera angles were categorised by the vertical angle and positions into a structured concept called Viewpoint. A Viewpoint serves to cluster a range of similar camera angles together. For this work, three viewpoints were established: Broadcast, Audience, and Overhead. These viewpoints are discussed in the following subsections.

Figure 3.1 - Tiled images of different viewpoints showing the diverse range of camera position and angles.

The broadcast viewpoint, shown in Figure 3.2, is most representative of official World Taekwondo footage and offers a standardised view of the competition area as per the World Taekwondo Event Operation Rules [13].

This viewpoint is prone to occlusion where occlusions refer to the obstruction or blocking of view from the side, where objects, individuals, or elements in the foreground prevent a clear line of sight to the area of interest. In the context of judging or observing a competition, lateral occlusions can make it difficult to accurately assess movements, positions, or interactions, particularly when the viewpoint is from ground level or at an angle where overlapping elements are more likely to occur. In Figure 3.2 the referee is partially blocking the red athlete's arm.

Figure 3.2 - Broadcast Viewpoint

Figure 3.3 shows an example of the audience viewpoint. This viewpoint is typically used by nations for their own performance analysis – in addition to scenarios where official competition recordings are not available. As illustrated in Figure 3.3 this is often from a slightly elevated position and can be recorded by any audience member.

Figure 3.3 - Audience Viewpoint

As illustrated in Figure 3.4 the overhead viewpoint is a direct or slightly angled view from above the competition area. This perspective reduces occlusions because it provides a clearer, unobstructed view of the entire scene, minimising the chances of objects or participants blocking one another from view. This is particularly useful for judges and audiences who need to review or analyse the bout, as it allows for better visibility and understanding of the competition dynamics. The overhead viewpoint has gained popularity since 2017, especially in situations where replays are required for accurate assessment of gam-jeom's or audience engagement.

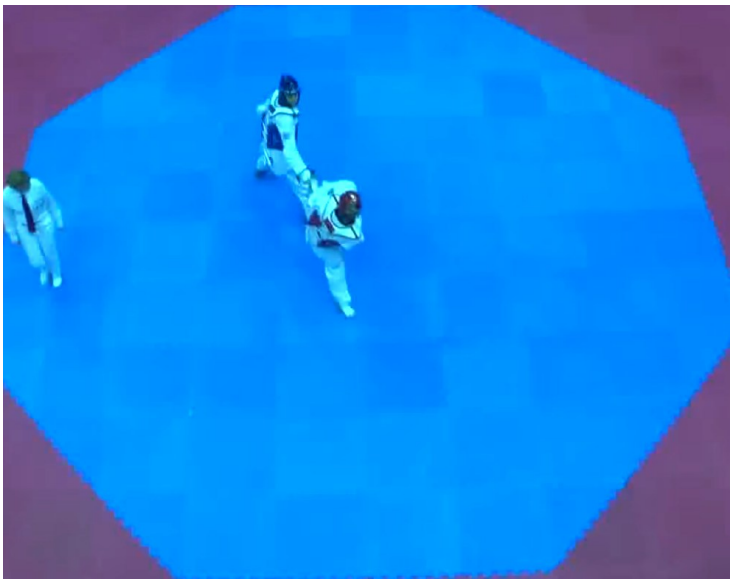


Figure 3.4 – Overhead Viewpoint

3.2.3 Framerate

As this work focused on individual image analysis to identify positions and actions, rather than temporal analysis such as distance over time, the issue of framerate was not deemed critical. Given that Taekwondo is a fast-paced combat sport, very low-framerate cameras were noted to potentially introduce artefacts, such as blurring believed to be caused by a short shutter speed. To mitigate this issue, only commonly accepted framerates, such as 24, 30 and 60 fps, were utilised. Videos with framerates above 60 were not considered, as the archive contained very few recordings of Taekwondo bouts where high-speed cameras were used.

3.2.5 Video Resolution

As illustrated in Figure 3.8, the quality of video footage has changed over time. In 2014, all recordings were captured in 540p resolution. By 2019, 540p accounted for only 44% of the footage, while 720p and 1080p resolutions collectively represented 51%. This trend has continued, resulting in the complete phasing out of 540p by 2023.

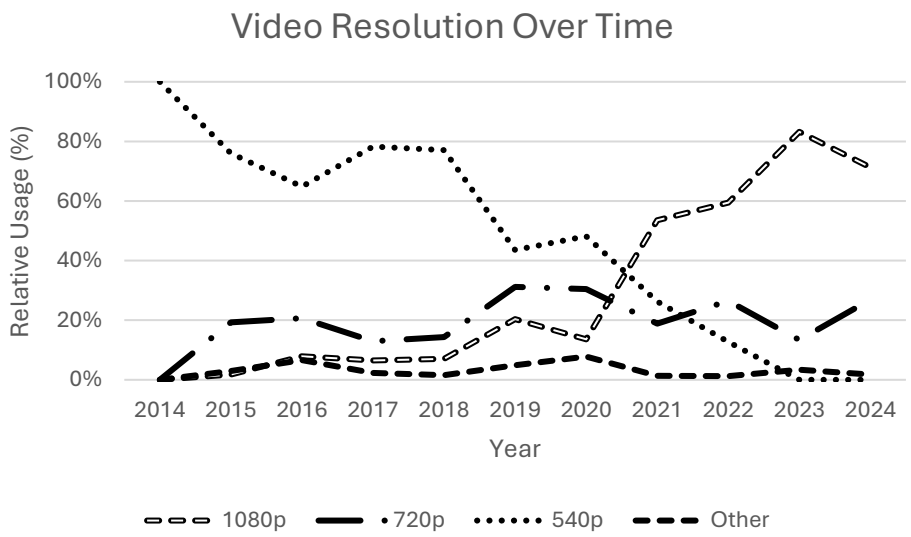


Figure 3.8 – Video resolution changes over time

Despite this progression, Figure 3.9 reveals that 540p still constitutes over 50% of the total footage historically. Given that the objective of this research study is to examine the validity of both manual annotation and CNNs when applied to historical footage, it is essential to ensure that samples from 540p, 720p, and 1080p resolutions are included in the dataset.

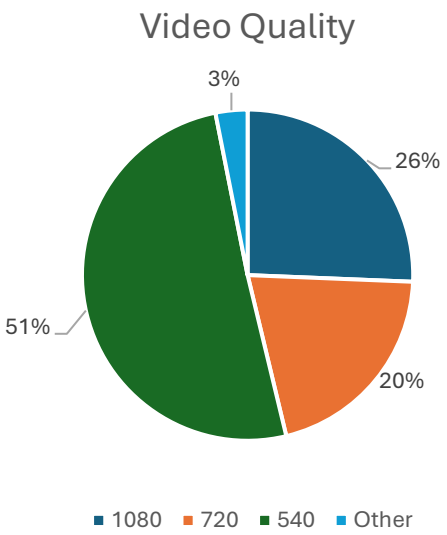


Figure 3.9 – Proportion of video resolutions (2014 - 2024).

As discussed in Chapter 2, Taekwondo frequently undergoes rule changes, which influence the way performance analysis is conducted. Typically, such analysis compares current data with that of the previous year. As data analysts and researchers often investigate the impacts of these rule changes - such as the measuring the incidence rate following a change in the head kick scoring method [86] the inclusion of footage from a broader range of years and varying quality levels ensures that the analysis addresses these scenarios.

3.2.6 Final Video Selection

In total, nine rounds were extracted from five unique international-level competitions, equally distributed across the three viewpoints (broadcast, n=3; audience, n=3; overhead, n=3) illustrated in Figure 3.5. The selection criteria ensured representation of 13 elite-level athletes at a consistent frame rate (25 fps), and varying resolutions (540, 720p, 1080p).

Figure 3.5 – Stills from each of the 9 rounds.

Full rounds were annotated (rather than random sampled frames) to:

- Capture movement diversity: Ensuring inclusion of all tactical scenarios (e.g., kicking phases, footwork transitions) that otherwise might be missed with random sampling.

- Account for variability: Differences between athletes influence kick dynamics and positional patterns [20], necessitating continuous tracking of various athletes to include their full tactical ability.
- Balance dataset scope: 35,913 annotated frames provided sufficient data depth while remaining feasible for manual annotation (vs. sparse random sampling across hundreds of rounds).

This approach prioritised comprehensive movement representation over volume of rounds, aligning with the research studies focus on positional validity across realistic competitive contexts.

3.3 Athlete Position Measurements

Athlete position measurements were captured through two complementary annotation approaches: (1) projected centre of mass (CoM) coordinates and (2) bounding box dimensions. These were supplemented by contextual measurements of kicking actions to account for movement-specific uncertainties. All digitisation was performed using CVAT [32], with every frame across the nine selected videos annotated to ensure comprehensive coverage of Taekwondo movement patterns and measurement scenarios. The methodologies for each measurement type and their reconstruction are detailed in the following sections.

3.3.1 Projected Centre of Mass

The projected centre of mass (CoM) represents the 2D approximation of an athlete's balance point during movement, mapped onto the competition mat plane, resulting in UV coordinates which can further be mapped into real world positional data on the mat. The midpoint between the athlete’s feet was used as the primary reference for the projected centre of mass, as this approach is widely adopted in literature [2], [22], [27], this midpoint serves as a kinematic proxy for athlete positioning similar to the output of pose estimation models within computer vision, while bounding boxes (Section 3.3.3) replicate object detection approaches and an approximation for the midpoint between athlete’s feet.

Each frame generally contained one of the following scenarios for each athlete:

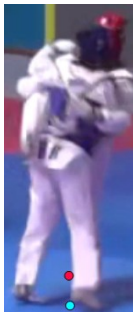
- **Both feet visible**: The midpoint between the athlete's feet was annotated directly Figure 3.6 - A.

- **One foot occluded**: The midpoint was estimated based on the visible foot and the expected stance, using contextual frames (before and after the occlusion) to inform the best guess (Figure 3.6 - B)

- **Both feet occluded**: The athlete's position was estimated using the frames immediately before and after, known as contextual frames, and the opponent's position as a reference (Figure 3.6 - C). These scenarios are illustrated in Figure 3.6, which provides examples of how annotations were handled in cases of partial occlusion or unusual positioning.

A – Both feet visible

B - One Foot Occluded



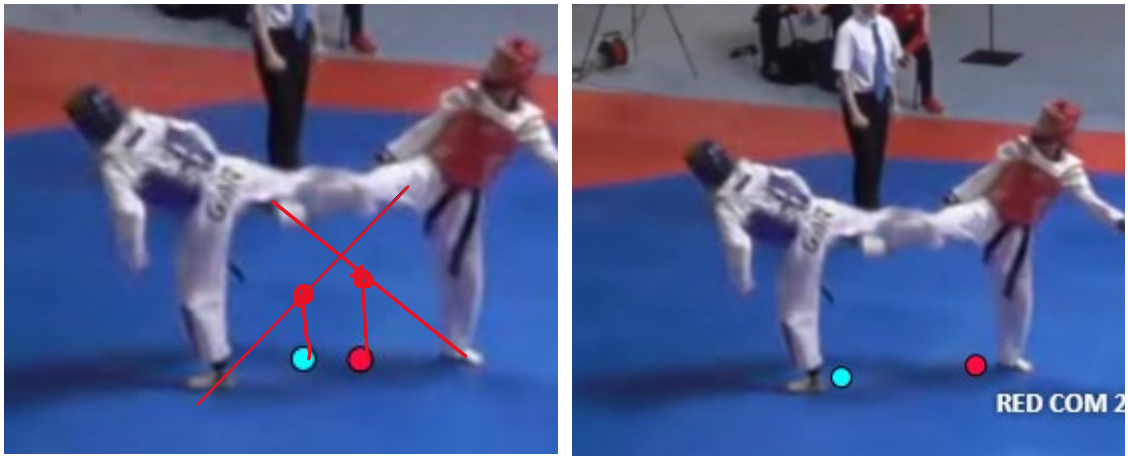
C – Both Feet Occluded

Figure 3.6 – Scenarios

The primary output of this annotation process was a set of UV coordinates for each athlete projected centre of mass in every frame.

3.3.2 Contextual Measurements

The annotation of kicking actions was included to their significant impact on measurement uncertainty during digitisation. When a foot leaves the ground (toe-off), both manual annotation and automated tracking face increased challenges in accurately determining athlete position, particularly when estimating either the midpoint between feet or the projected centre of mass (CoM). As shown in Figure 3.7 - A and B, these scenarios produce greater positional ambiguity compared to when the athlete has both feet on the ground.



A – Close to the midpoint B – Close to 'centre of mass'

Figure 3.7 - Annotation error during kicks

To capture this uncertainty, the start (toe-off) and end (toe-on) points of each kick were annotated, following established conventions in literature [2], [15], [87]. The contact moment with the opponent was not recorded, as the study focused on the complete kicking motion's positional characteristics rather than the interpersonal dynamics of the athletes. Figure 3.8 demonstrates the typical movement sequence (B, C, D, C, B, A) during a Taekwondo kick, highlighting the phases where positional estimation becomes most challenging.

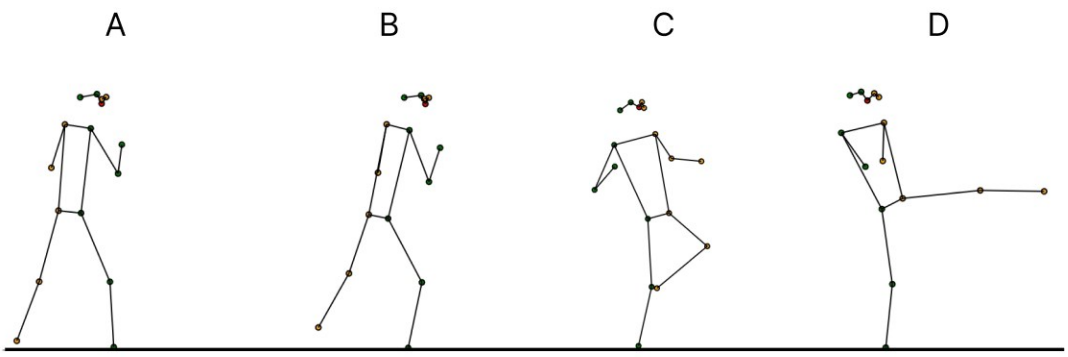


Figure 3.8 – Pose of a Taekwondo Kick

3.3.3 Bounding Box

For each frame of every video, bounding boxes were measured to capture the position and extent of each athlete. This process involved drawing a rough rectangle around both athletes and refining the anchor points until the box precisely overlapped each athlete. The bounding box approach was selected based on its established use in automated athlete tracking research across various sports [88], [89]. The bounding box data also serves as a foundation for training a CNN bounding box model and understanding domain-specific challenges, such as occlusion in future chapters.

Figure 3.9 shows each frame contained one of the following scenarios for each athlete:

- **No occlusion**

  The bounding box was annotated directly, capturing the full extent of the athlete's visible body. In these cases, the bounding boxes for both athletes were fully apart, with no overlap (

  Figure 3.9 - A).

- **Partial occlusion**

  In cases where only part of the athlete (e.g., the head or one limb) was visible due to occlusion by the opponent or referee, the bounding box was estimated using contextual frames (i.e., frames immediately before and after the occlusion) to infer the athlete's likely position. In these cases, the bounding box generally overlapped but never fully (

  Figure 3.9 - B).

- **Full occlusion**

  When an athlete was completely occluded, their position was estimated based on contextual frames and the opponent's position. In these cases, the bounding boxes often fully overlapped, with one box contained within the other (

  Figure 3.9 - C).

A – No occlusion B – Partial occlusion C – Full occlusion

Figure 3.9 – Examples of occlusion

The primary output of this process was a set of bounding box coordinates for each athlete in every frame. These coordinates were used to derive additional metrics, such as the midpoint of the bottom edge of the bounding box, which closely aligns with the athlete's

stance on the canvas. This midpoint provides a reliable positional metric for tracking lateral and vertical movement, minimising changes introduced by posture or limb extension.

3.3.4 Reconstruction

Athlete positions were recorded in two coordinate systems: the image coordinate system (UV) and the world coordinate system (XYZ). The UV coordinates represent athlete positions within the 2D plane of each video frame, while the XYZ coordinates provide a real-world spatial representation after reconstruction.

The importance of separating these two coordinate systems lies in understanding and addressing the distinct sources of error introduced at each stage. UV coordinates are subject to error from manual annotations or automated methods, while XYZ coordinates introduce additional error during the reconstruction process due to factors such as camera calibration and out-of-plane error. By keeping these systems separate, it is possible to better assess the impact of these errors on the validity of using CNNs, ensuring more reliable and accurate athlete position measurements.

3.3.5 2D Direct Linear Transformation (DLT)

To map image coordinates (UV) to real-world coordinates (XYZ), the 2D Direct Linear Transformation (DLT) method was employed, as discussed in the literature review (Section 2.5) [81]. This method was selected over other reconstruction techniques due to its ability to handle perspective distortions caused by varying camera angles and positions on historical footage, without the use of an existing calibration matrix such as a checkerboard for intrinsics, which is critical for accurately analysing Taekwondo competition footage.

In this study, the eight vertices of the Taekwondo competition area (labelled A1 through A8 in Figure 3.10) were used as reference points. These points were chosen over a minimal set of four to better account for perspective distortions and measurement errors across the entire competition area. By identifying these points in the video footage, a transformation matrix was computed and applied to map pixel coordinates to real-world coordinates.

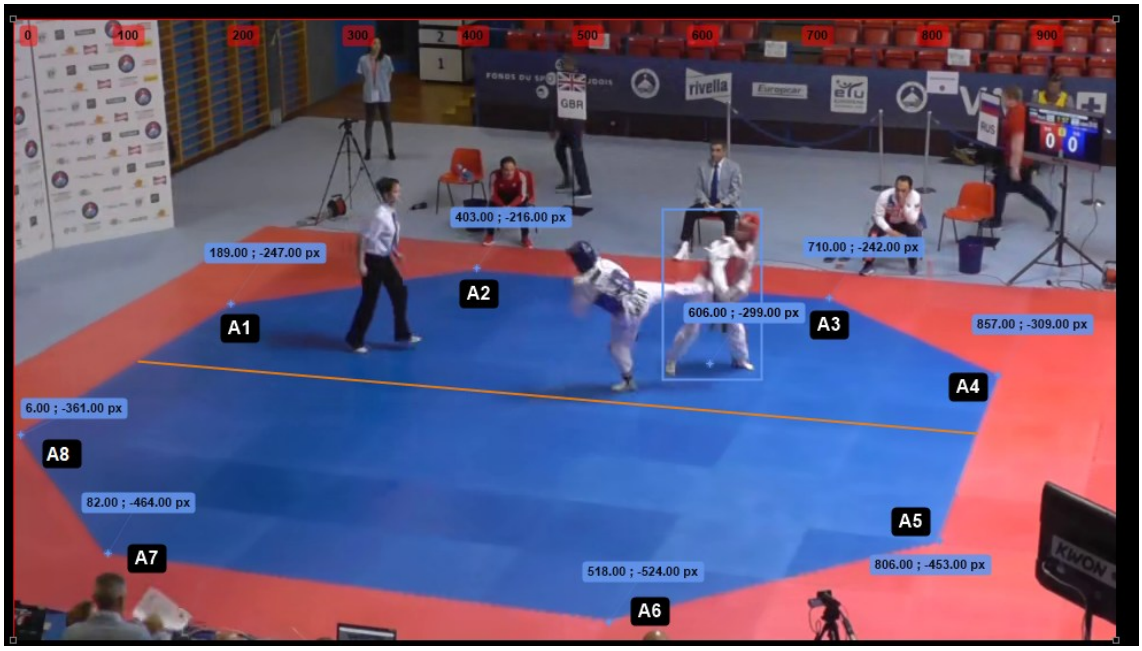


Figure 3.10 - Annotation Calibration Points

3.3.6 Summary

A total of 35,913 frames were annotated, with each frame recording the bounding box for both athletes, the projected centre of mass position for each athlete, and whether each athlete was kicking at that time. In total this resulted in 143,652 bounding box coordinates, 71,826 CoM positions, and 596 recorded kicks.

3.4 Conclusion

The chapter addressed the research studies first objective: to create a representative dataset that captures the different scenarios of Taekwondo competition footage. The objective has been achieved through the careful selection and annotation of video footage, ensuring diversity in viewpoints, athlete representation, video resolution and tactical scenarios, additionally it allows for the domain of Taekwondo to be understood from a data driven perspective, alongside validation of manual error and to enable two separate machine learning models to be explored.

Due to the size of the archive, it was impractical to utilise all available data. Manual annotation and digitisation of such a large dataset was determined to be infeasible within the constraints of this research. It was estimated that each round takes approximately four hours to annotate, with each video containing a minimum of two rounds, resulting in over 35 years of total manual annotation time required.

The dataset comprises of 9 videos extracted from 5 unique international-level competitions, equally distributed across three viewpoints: broadcast, audience, and overhead. Including 13 unique athletes, 3 unique resolutions (540p, 720p, 1080p) and a consistent framerate of 25fps, providing a comprehensive foundation to address the aim of the research study: To explore the validity of CNNs in Taekwondo.

These measurements enable the next stages of the research study: (2) a deeper understanding of athlete movement patterns (Chapter 4), (3) the quantification of error in manual annotation (Chapters 5 and 6), and (4) the validation of automated tracking methods (Chapter 7).

Chapter 4 – Understanding the domain

4.1 Introduction

Building on the competition dataset from Chapter 3, this chapter systematically analyses Taekwondo bout footage to identify patterns and challenges in measuring athlete interactions. While prior chapters established that spatial annotation data can be extracted from real-world footage, this chapter maps the domain of Taekwondo footage into various scenarios in which it is possible to evaluate under what conditions these measurements remain reliable and valid, producing a foundation for evaluating both manual (Chapters 5 – 6) and automated (Chapters 7 – 9) methods in future chapters.

In this chapter the following metrics are explored:

1. **Viewpoint**: Quantifying camera angles to stratify measurement difficulty
2. **Bearing**: Capturing the position of the athlete in relation to each other
3. **Containment**: Mapping how athlete bounding boxes interact during attacks or blocks

In addition to these metrics, distance and whether the athletes are performing a kick are also integrated into the analysis for contextual information. The output is a data-driven classification of Taekwondo annotation scenarios; these classifications may highlight certain scenarios where manual annotation or CNN reliability might struggle. This framework enables targeted reliability testing in later chapters, moving beyond binary reliable or not reliable claims to scenario-specific performance evaluation.

4.2 Methodology

4.2.1 Data preparation

The dataset used in this analysis comprises 35,913 continuous frames from 9 videos at 25fps, as described in Chapter 3. Each frame was annotated with bounding boxes, projected centre of mass (CoM) positions, and kick/no-kick labels. All procedures were approved by Sheffield Hallam University Research Ethics Board (ER39637393).

4.2.2 Viewpoint

During the annotation process the viewpoint (the position of the camera relative to the athletes) had an influence on the reliability of annotated measurements in Taekwondo. Viewpoint was examined as a potential source of measurement bias, as camera angles influence the perceived scale and overlap of athletes (e.g., overhead views minimise occlusion, whereas broadcast views increased occlusion).

This analysis aimed to identify scenarios where measurement reliability could be compromised, such as high containment during kicks from a lateral viewpoint.

4.2.3 Bearing

Bearing was defined as the angle between the two athletes relative to the camera, providing a precise measure of their relative positioning to one another. Specifically, the bearing was calculated as the direction of the blue athlete from the red athlete's position as illustrated in Figure 4.1, with the camera serving as the reference point. This metric was used to analyse the relationships between the athletes, particularly in scenarios involving movements and occlusion. This metric was chosen as depending on the athletes' relative position the annotation scenario may become more difficult due to various factors such as athlete occlusion.

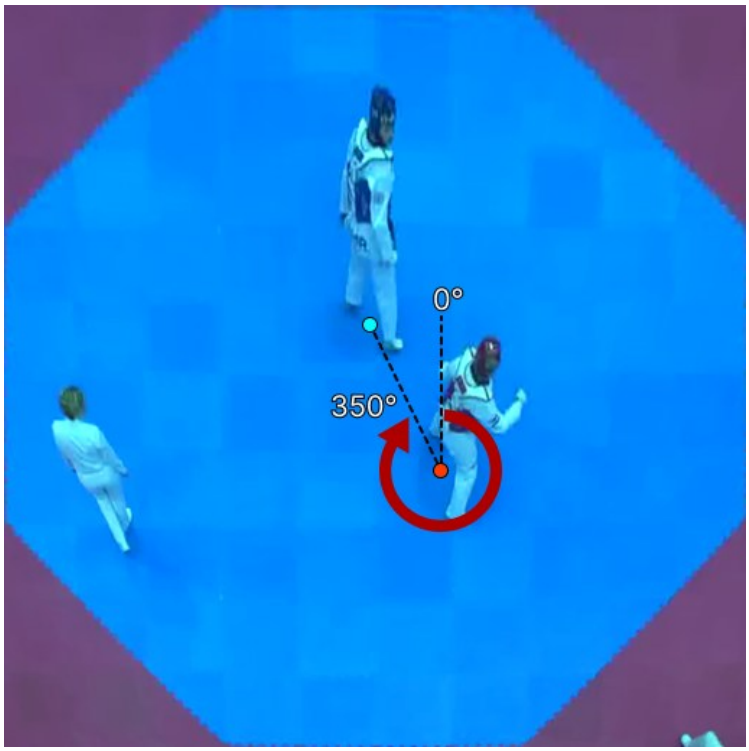


Figure 4.1 - Calculation of bearing

The calculation of bearing was based on the transformed coordinates of the athletes' projected centre of mass (CoM) positions in metres. Where the origin is the bottom left of the viewpoint. The bearing was calculated for every frame. This continuous measure allowed the analysis of how athletes positioned themselves relative to each other over time.

Bearing is used for analysing the duration and frequency of specific positional scenarios. A bearing of 0 degrees indicated that the red athlete was directly in front of the blue athlete, potentially occluding them from the camera's view. Conversely, a bearing of 180 degrees indicated that the blue athlete was directly in front of the red athlete giving two similar but differing annotation scenarios. Other angles represented varying degrees of side-by-side positioning, providing a comprehensive view of the athletes' movement.

4.2.4 Containment and Overlap Analysis

During the data collection process in chapter 3, it was observed that occlusion scenarios in Taekwondo vary widely in complexity, not just based on the athletes relative positioning. The literature review chapter also highlights this is not an uncommon issue, previous studies attempted to categorise into ‘occluded’ and ‘not-occluded’ this is not sufficiently detailed for Taekwondo bouts. For instance, where slight occlusion occurs but the annotation remains straightforward as illustrated in Figure 4.2, which shows an example of a partially occluded athlete where the position annotation is still feasible.

Figure 4.2 - Partially occluded athlete annotation

In other scenarios, occlusion makes annotation unreliable as illustrated in Figure 4.3. To investigate occlusion and its impact on annotation accuracy, two primary metrics were explored: Intersection over Union (IoU) and Containment (A in B).

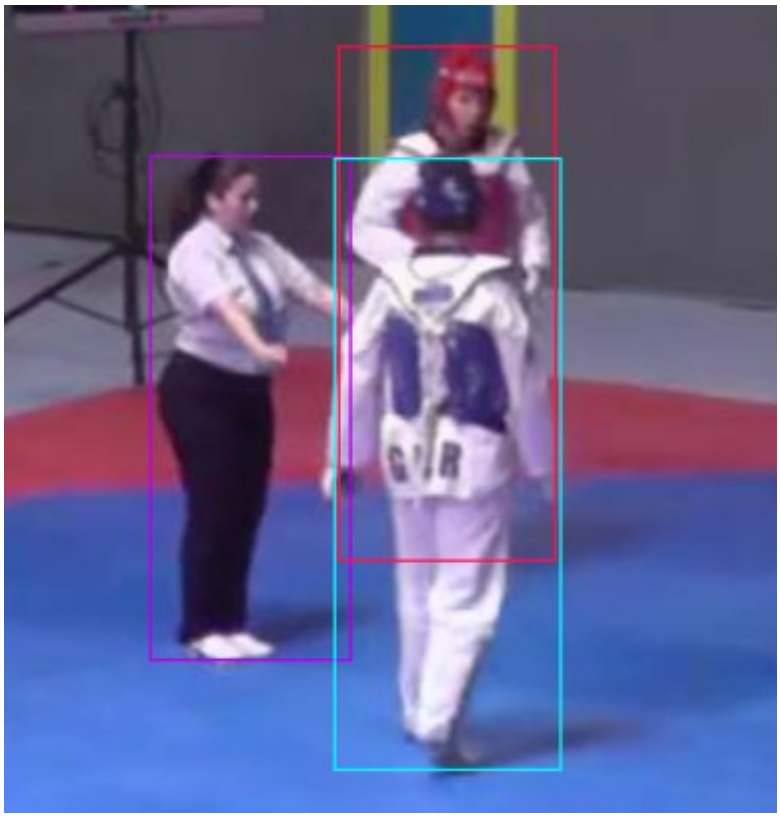


Figure 4.3 - Major occluded athlete annotation

Bounding Box Overlap (IoU)

As discovered in chapter 2, intersection over Union (IoU) is a widely used metric in object detection and annotation tasks. It measures the overlap between two bounding boxes by dividing the area of their intersection by the area of their union.

While IoU is effective for comparing the similarity of bounding boxes in terms of size and location, it has limitations when applied to measuring occlusion scenarios. Specifically, IoU often yields low percentages even when one bounding box is almost entirely contained within another. This occurs as IoU is sensitive to the relative sizes of the bounding boxes.

In Taekwondo, athletes closer to the camera have larger bounding boxes, while those further away have smaller bounding boxes. As a result, even when one athlete is fully or nearly fully occluded by another, the IoU value may remain low due to the disparity in box sizes as illustrated by Figure 4.4 where the IoU figure is only 44.6% even despite the red athlete being fully occluded. The metric of IoU is best used when comparing a ground truth box to an annotated box as this allows you to measure the exact error between two boxes.

Figure 4.4 – Example of occlusion

Containment (A in B)

Given the limitations of IoU in this context, a containment-based approach was adopted to better quantify occlusion. Containment measures how much of one athlete's bounding box (A) is contained within another athlete's bounding box (B) where box A is always the smaller box (or the athlete furthest from the camera) and box B is always the larger box (the athlete closest). This is particularly useful for analysing occlusion scenarios, as it directly addresses the issue of one athlete being obscured by another. The containment ratio is calculated as:

$$Containment = \frac{A \cap B}{A}$$

This metric provides a more intuitive measure of occlusion, as it focuses on the proportion of the occluded athlete's bounding box that is hidden by the occluding athlete. For example, if Athlete A's bounding box is entirely contained within Athlete B's bounding box, the containment ratio will be 100%, indicating full occlusion. This contrasts with Intersection over Union (IoU), which may yield substantially lower values (e.g., 44.6% in Figure 4.4) due to differences in bounding box size.

4.3 Results

4.3.1 Containment

Containment metrics (A in B, Section 4.2.4) were analysed in relation to the positional categories (front, side, back) defined in Section 4.3.2. This provided a quantitative measure of occlusion across different athlete configurations and viewpoints.

As shown in Figure 4.5, containment values showed frequent spikes throughout bouts, these spikes correlated with movements such as kicks or forward lunges, in essence where athletes encroached each other's interpersonal distance. The frequent occurrence of these events indicating tactical actions highlights the need to stratify containment, as a spike in containment can indicate either a difficult annotation scenario such as when an athlete is occluding another, or where a kick or other tactical move has occurred.

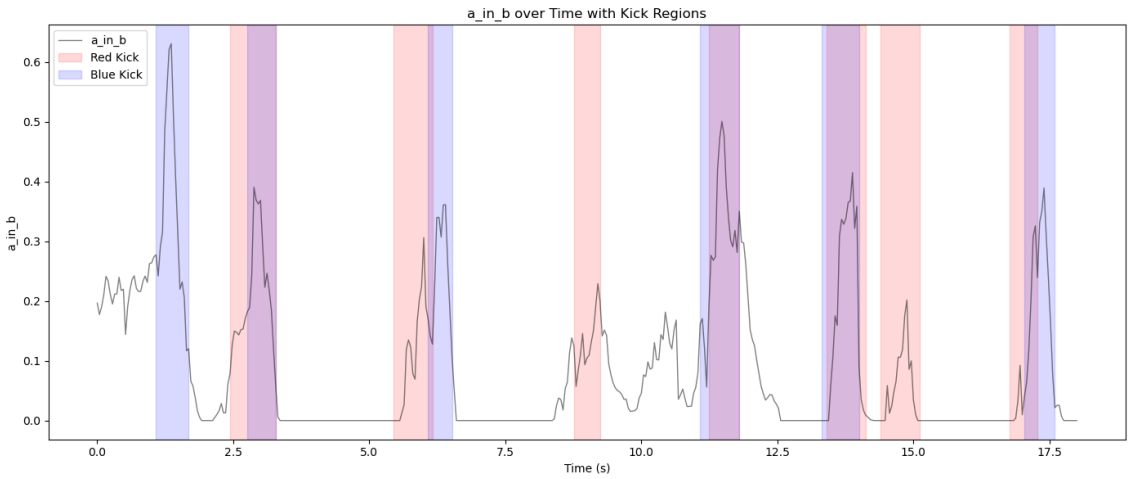


Figure 4.5 - Containment vs Time

Since increases in containment align with tactical actions, this relationship is used to generate a histogram representing the level of activity within a round (Figure 4.6), this figure shows that approximately 68% of the time the athletes are inactive (When containment < 5%).

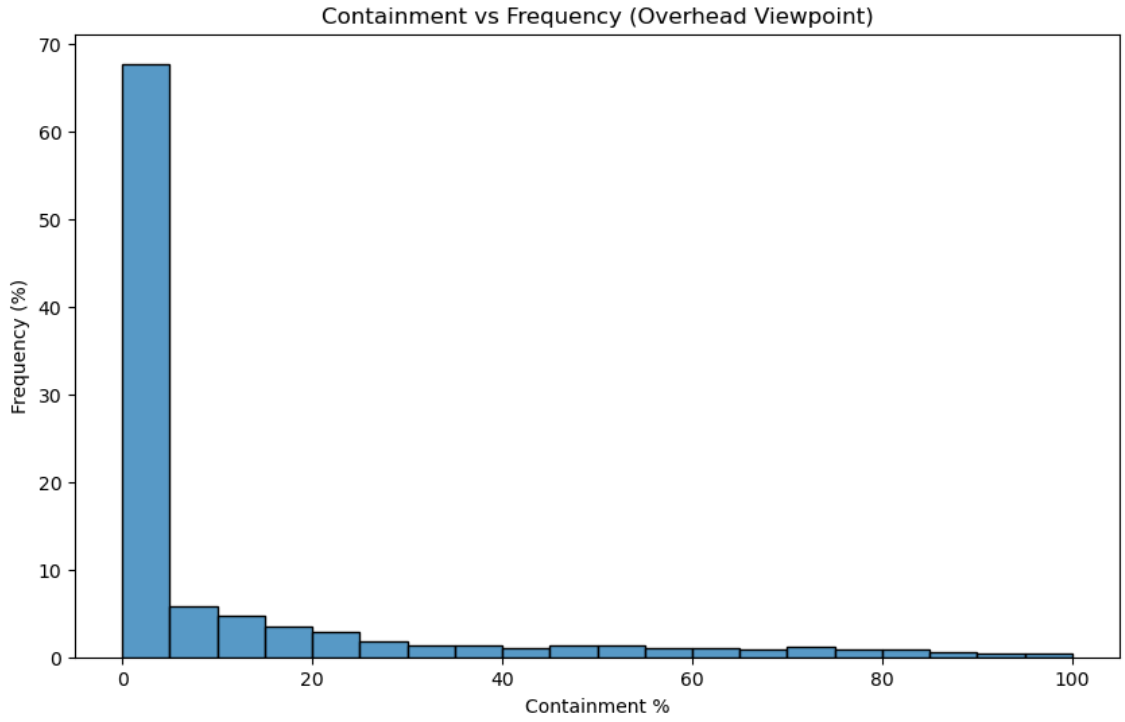


Figure 4.6 - Containment vs Frequency by Viewpoint

To simplify further analysis, containment values were discretised into five categories represented for each viewpoint in Figure 4.7:

- <1 % - No Containment
- 1 - 25% - Minimal Containment
- 25 - 50% - Partial Containment
- 50 - 75% - Moderate Containment
- 75 - 100% - Major Containment

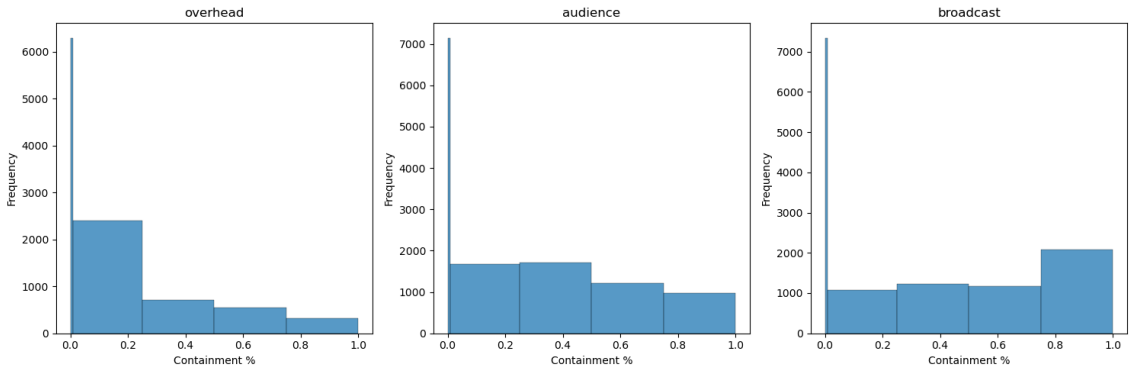


Figure 4.7 - Containment histogram

4.3.1 Viewpoint

Viewpoint Distortion

For each viewpoint the Taekwondo area was recorded horizontally (U) and vertically (V) in pixels. From this several metrics were calculated: The absolute and relative dimensions were measured in pixels (UV) and as a percentage of the total resolution. The minimum detectable difference (MDD) representing the smallest measurable change in athlete position. The aspect ratio (U/V) was also computed to assess dimensional balance, with values closer to 1 indicating more equal representation of each axis. These dimensions are illustrated in Figures 4.8, 4.9 and 4.10 with the results in Table 4.1 below.

Table 4.1 – Viewpoint Distortion Results

Viewpoint	Absolute Size (px)		Mat Coverage		MDD (cm)		Ratio	Resolution
	U	V	U	V	X	Y		
Overhead	762	792	40%	73%	1.05	1.01	0.96	1080p
Broadcast	1280	308	100%	42%	0.63	2.60	4.16	720p
Audience	910	187	95%	35%	0.88	4.28	4.86	540p

The overhead viewpoint showed the most balanced dimensional representation, with U and V covering 40% and 73% of the resolution respectively (Figure 4.8). This configuration produced an aspect ratio of 0.96 and the smallest MDD values (1.05 cm horizontal, 1.01 cm vertical), indicating high sensitivity to positional changes. In contrast, the broadcast viewpoint (Figure 4.10) showed significant dimensional imbalance, with full horizontal coverage (100% U) but only 42% vertical coverage, resulting in an extreme aspect ratio of 4.16. This configuration created substantial disparity between horizontal (0.63 cm) and vertical (2.60 cm) MDD values.

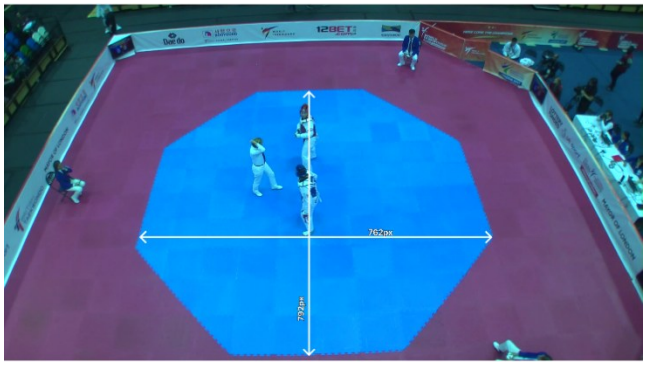


Figure 4.8 - Overhead viewpoint U, V measurement (state U and V on image)

Figure 4.9 - Audience viewpoint, U, V measurement

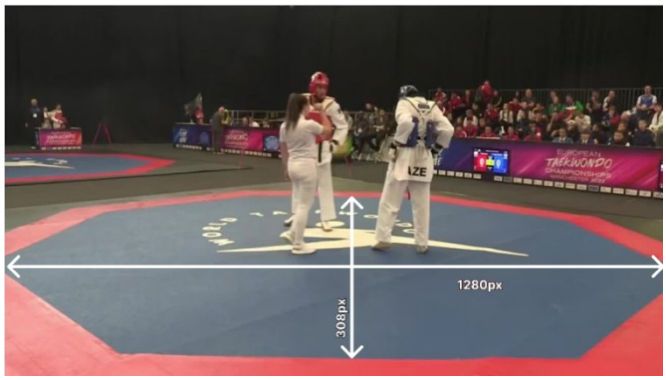


Figure 4.10 - Broadcast viewpoint, U, V measurement

Variations in measurement reliability were observed across different camera angles. The overhead perspective was found to demonstrate measurement consistency, characterised by minimal perspective distortion and reduced occlusion frequency. This viewpoint provided uniform scaling of athlete positions, which allowed for more accurate distance calculations. The elevated angle presented limits in depth perception during dynamic techniques such as kicks.

Viewpoint Occlusion

To better understand how occlusion risk manifests across different camera perspectives, containment levels were stratified by viewpoint. As described in Section 4.2.4, containment (A in B) offers a context-sensitive measure of visual obstruction between athletes. In this analysis, containment frequencies were normalised by the total frame count per viewpoint to allow percentage-based comparisons. Containment was then categorised into four levels as outlined previously, with 'no containment' being excluded:

- <1 % - No Containment
- 1 - 25% - Minimal Containment
- 25 - 50% - Partial Containment
- 50 - 75% - Moderate Containment
- 75 - 100% - Major Containment

The broadcast viewpoint exhibited a balanced distribution of containment levels, with a slight skew toward major containment (Figure 4.11). This suggests that while broadcast footage can capture a wide variety of occlusion intensities, it disproportionately includes frames where athletes are tightly overlapped. This may reflect the lateral camera position which increases the likelihood of one athlete obstructing the other along the horizontal axis. As a result, this viewpoint introduces moderate risk to measurement reliability, particularly during close engagements such as kicks.

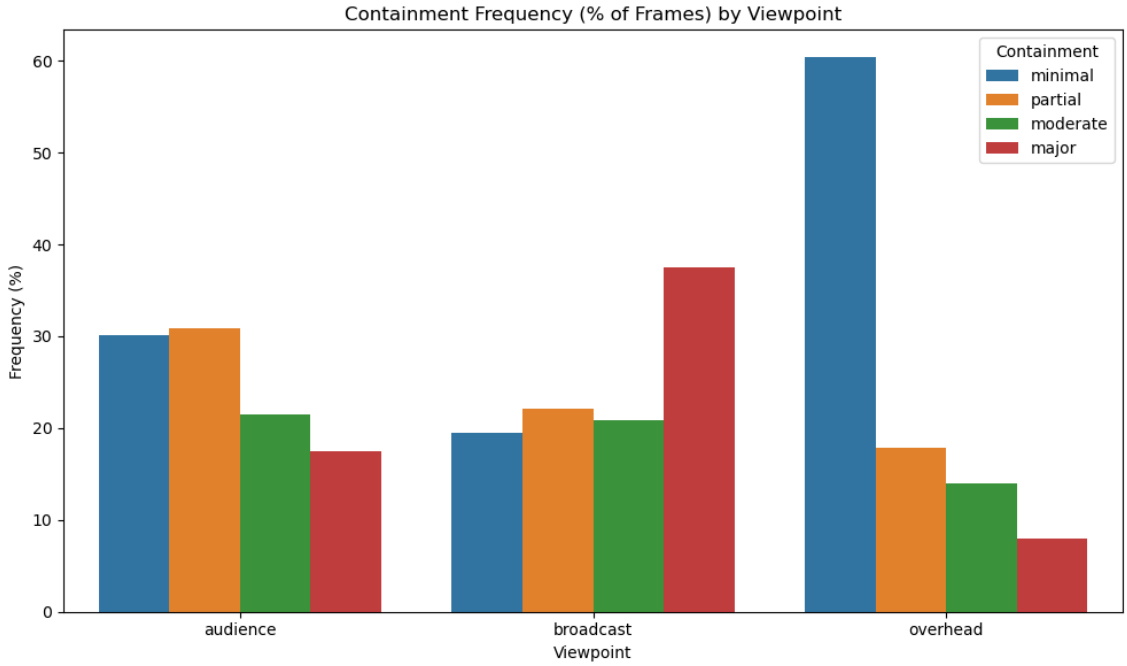


Figure 4.11 - Containment frequency by viewpoint

The audience viewpoint showed a shift toward lower containment levels, with the highest frequencies observed in the partial and minimal categories. This suggests that audience-angle can capture the athletes where neither athlete fully occludes the other. This low major containment stems from the camera's offset perspective, which tends to compress vertical spatial information and reduce instances of direct overlap.

The overhead viewpoint demonstrated a distinct pattern where minimal containment accounted for most frames, followed by sharp drop-offs in higher containment categories. This confirms earlier findings that overhead views minimise occlusion due to the top-down perspective and relatively even scaling of athlete bounding boxes. The rarity of major containment in this viewpoint reinforces its suitability for accurate spatial measurement, particularly in dynamic scenarios.

4.3.2 Bearing

The relative bearing between athletes, defined as the angle of the blue athlete relative to the red athlete with the camera as the reference point, was analysed to assess how

positioning influences measurement reliability. Continuous bearing data revealed patterns in athlete positioning, such as orientation shifts during offensive or defensive actions, while also establishing a foundation for evaluating occlusion risks through containment analysis (Section 4.3.3).

Figure 4.12 illustrates the temporal evolution of bearing angles during a representative bout. The athletes predominantly occupied the angles between 220° and 280°, with an abrupt positional switch at 60 seconds into the round for approximately 400 frames (16 seconds), indicating tactical repositioning. This dynamic is further clarified in Figure 4.13 which shows a polar plot representation where the blue athlete's trajectory shifts from the left to the right hemisphere before returning, highlighting transient dominance phases.

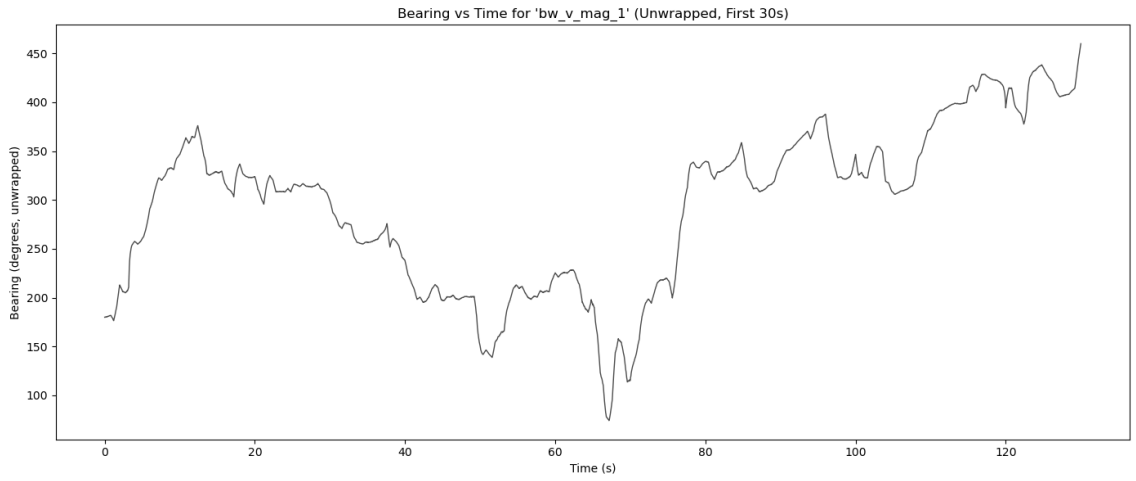


Figure 4.12 - Bearing vs Frame (Continuous)

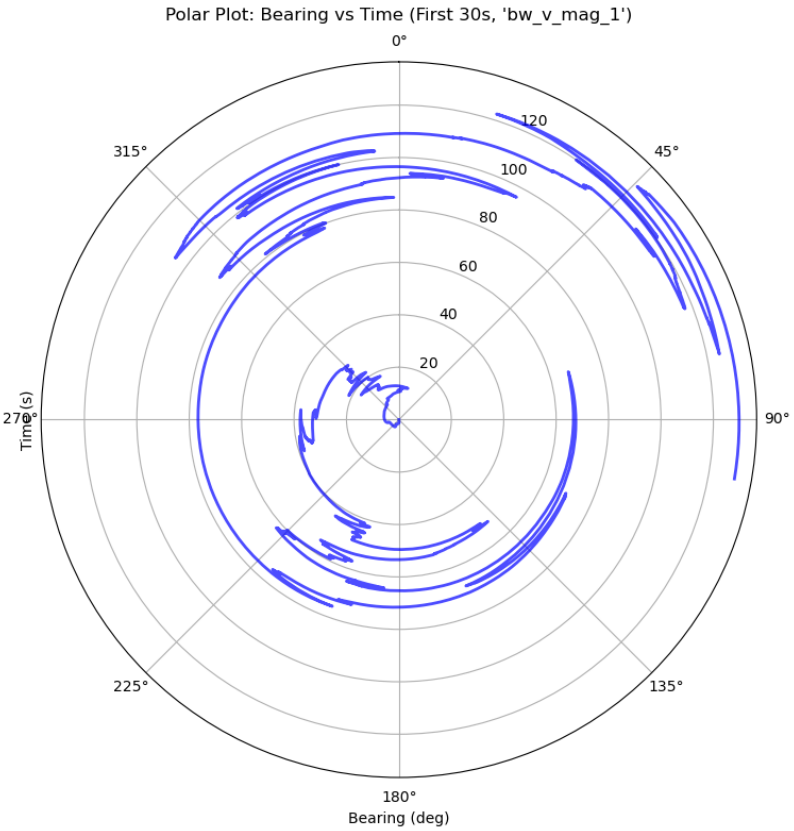


Figure 4.13 - Bearing vs Frame (Continuous, Polar)

Athlete Orientation

Outlined in section 4.2.3, continuous bearing data was then used to qualitative classify each athlete into front, side, or back scenarios. To quantify this front, side and back bias metric for each of the viewpoints a set of tiled images were generated, where unique frames were extracted at 15° intervals alongside 10% containment increments. Frames not meeting proximity thresholds (±2.5°, ±5% containment) were excluded, as shown in Figure 4.14.

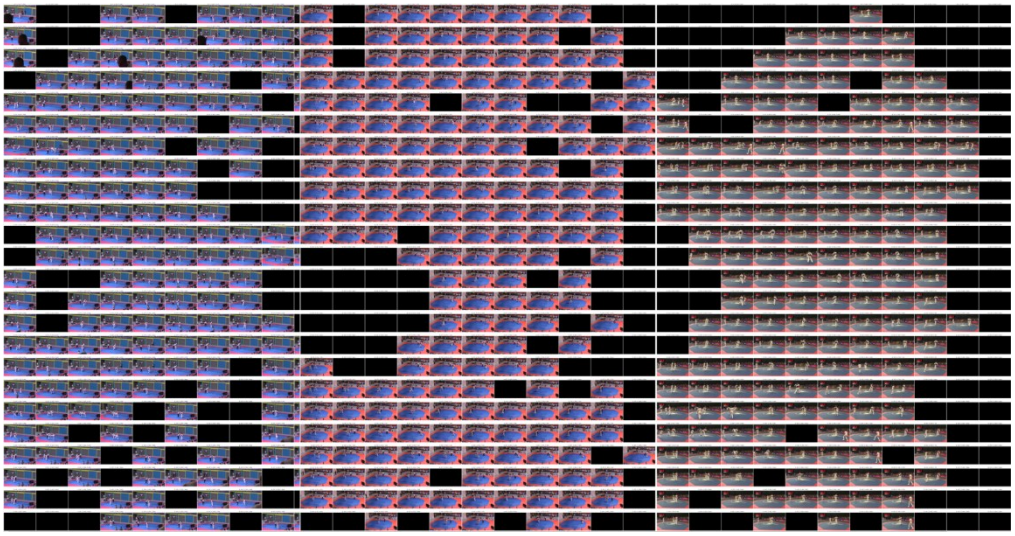


Figure 4.14 - Broadcast, audience and overhead views showing bearing and overlap.

The results are presenting in Table 4.2. From the manual classification of the tiles, it was discovered that due to perspective distortion, often what would be considered a 'back' view in an overhead would be considered a 'side' view when viewed from a broadcast viewpoint. This was based on the visible orientation of the athletes' torsos and feet relative to the camera, as well as the degree of lateral separation between the athletes' projected centres of mass. As such each viewpoint presented specific angular thresholds. These differences are illustrated in figure 4.15.

Table 4.2 – Viewpoint based classification of orientation

Viewpoint	**Back °**		**Front °**		**Side °**	
Overhead	315	045	135	225	045	135
					225	315
Audience	300	060	120	240	060	120
					240	300
Broadcast	285	075	120	240	075	120
					240	285

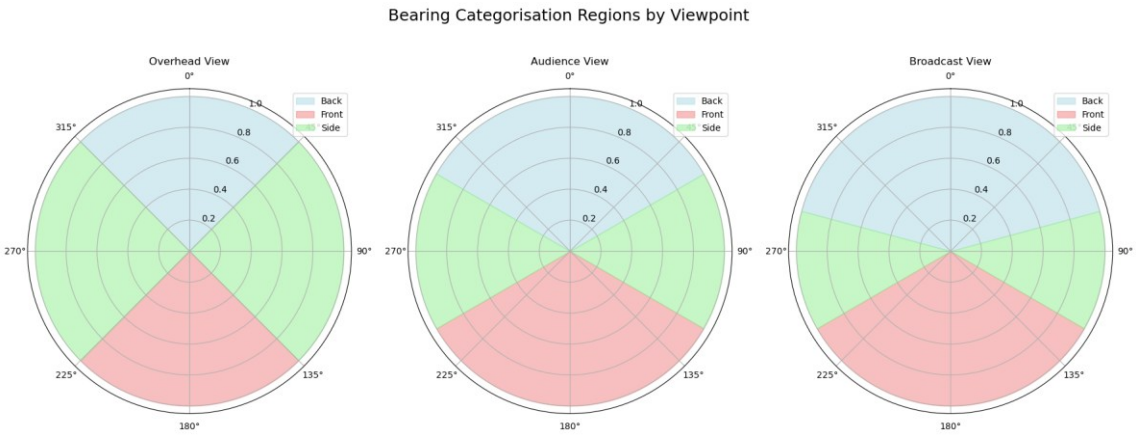


Figure 4.15 – Front Back and Side views by their Viewpoint

These discrete classifications mitigate the limitations of continuous bearing metrics by reducing sensitivity to small angular fluctuations and perspective-induced distortion. By grouping athlete configurations into front, side, and back categories, occlusion risk can be stratified more robustly, as front and back scenarios are consistently associated with higher containment. This categorical framework also standardises annotation decisions across viewpoints, ensuring that similar interaction scenarios are treated consistently despite differences in camera angle.

As shown in Figure 4.16, athletes occupied a side alignment in 67% of frames, with behind and front positions accounting for 18% and 14%, respectively. This aligns with the tactical orientation of the sport where athletes maintain their starting side-on stances.

The lower percentages of time spent in front and behind configurations support the earlier observation that these scenarios are less frequent but critical as they often coinciding with occlusion-prone interactions. These front/back alignments are strongly correlated with increased containment values (Section 4.3.3), making them significant contributors to annotation difficulty despite their limited duration.

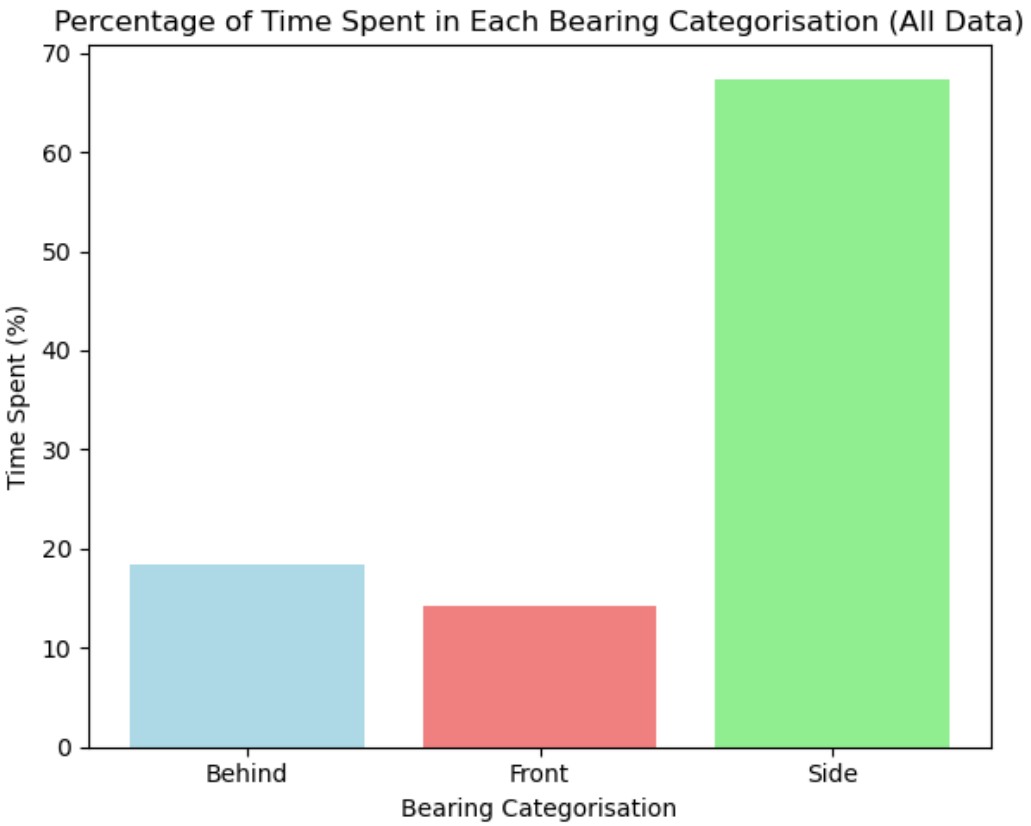


Figure 4.16 - Time spent per relative position

4.4 Established Framework

The combination of bearing, containment, and viewpoint analyses enables the development of a structured framework to classify Taekwondo scenarios. This framework identifies high-risk conditions where manual or automated annotation methods may struggle, moving beyond binary reliability judgments to context-specific evaluation. Three primary dimensions define the framework:

Viewpoint (Overhead, Broadcast, Audience):

Camera perspective significantly influences measurement validity, as demonstrated by disparities in containment distributions (Section 4.3.3) and minimum detectable differences (Section 4.3.4). The overhead viewpoint, with its balanced aspect ratio (U/V: 0.96) and low occlusion frequency, represents the most reliable scenario. In contrast, broadcast footage experiences severe perspective distortion (U/V: 4.16), highlighting occlusion in front/back configurations.

Bearing Position (Front/Side/Back):

Discrete bearing categories (Section 4.3.2) stratify athlete position and correlates with occlusion risk. Front/back alignments contain a naturally high reliability risk, especially so for the back athlete (>50% in 68% of broadcast cases), while side and front alignments often pose a smaller reliability risk.

Containment Severity (Minimal to Major):

Containment levels (Section 4.3.3) quantify occlusion intensity. Major containment (75 – 100%) occurs in 22% of front/back scenarios but only 3% of side alignments, suggesting annotation reliability degrades during close distances.

4.5 Discussion

This chapter systematically analysed Taekwondo athlete interactions through viewpoint, bearing and containment. The results highlight the challenges for annotation reliability, which directly inform the evaluation of manual and automated methods in later chapters. Each unique combination of these three parameters led to various high-risk scenarios where the expected reliability would be low, such as the broadcast viewpoint, back position and major containment, and low-risk scenarios where reliability would expect to be high for example in the overhead viewpoint, side position and minimal containment.

4.5.1 Occlusion

The data reveals that major containment (>75%) occurs in 22% of front/back configurations, this suggests manual and automatic annotators may struggle to track athlete positions in these scenarios. This aligns with broader literature on occlusion in sports analytics, where occlusion is known to reduce tracking reliability [90]. Research in both basketball and boxing has shown that side or broadcast camera angles often lead to high occlusion rates, especially during close interactions, while overhead perspectives drastically reduce occlusion issues [85], [91]. These findings reinforce the observation that the overhead viewpoint, with balanced dimensions and lower occlusion rates, provides a more reliable angle for athlete tracking and annotation tasks. From this, it is hypothesised that annotation error rates will rise substantially in high-containment scenarios, particularly when athletes are aligned front-to-back relative to the camera.

The overhead viewpoint was the most reliable perspective, with balanced dimensions (U/V ratio: 0.96) and lower occlusion rates (12% major containment). In contrast, broadcast footage (U/V ratio: 4.16) exhibited severe perspective distortion, which led to an occlusion of 68% of front/back cases (Section 4.3.4). This suggests that annotation on

overhead footage will outperform those using broadcast angles, as the latter will introduce compounding errors from occlusion, calibration and perspective bias.

4.5.2 Measurement challenges during dynamic actions

During kicks a 26% reduction of distance was observed (from 2.3 m to 1.7 m), which additionally correlated with an increased occlusion risk. Bearing angles remained stable across kick phases, indicating that kicks occurred in all positional contexts: front, side, or back. The short bursting nature of the kicks suggests that annotation systems, manual or automated must be able to capture and measure rapid positional changes during dynamic actions.

The consistency of bearing angles further suggests that camera perspective, rather than relative athlete orientation, is the primary driver of occlusion. This is consistent with biomechanics research showing that certain viewing angles introduce substantial measurement errors, up to 2.5 times greater when compared to perpendicular or overhead views [92]. This supports the argument that viewpoint distortion, particularly in broadcast footage, can significantly degrade the accuracy of distance calculations derived from video.

4.5.3 Evaluating Annotation Reliability

The framework developed in Section 4.3.6 enables the stratification of annotation scenarios. This stratification offers a structured way to assess where manual or automated methods may have higher error in later chapters.

Manual annotation will be explored in Chapter 5, where inter-rater reliability is tested across containment levels. Prior research shows that when scenarios are clearly defined and visual features are unobstructed, inter-rater agreement on athlete kinematics can be very high [93]. This supports the hypothesis that minimal-containment frames, which

make up 68% of the dataset, is likely to give reliable annotations suggesting on the surface that annotation of Taekwondo bouts is accurate. During the high-containment or dynamic sequences, where the data is most valuable, it is expected to have more disagreement due to occlusion.

4.6 Conclusion

By quantifying occlusion risks, viewpoint biases, and dynamic interactions, this chapter moves beyond binary claims of reliability to a scenario-based framework for Taekwondo. The findings from this chapter suggest that annotation accuracy is highly context-dependent, influenced by camera angle, athlete alignment and position.

These insights will guide the validity testing in later chapters, ensuring that manual and automated methods are evaluated under realistic conditions. The contextual dependency of annotation reliability is well-supported by studies in sports analytics, which highlight the effects of viewpoint [92] occlusion [90], [91], and scenario clarity [93]. The findings of this chapter align closely with these patterns and extend them specifically to Taekwondo using quantifiable containment and positional metrics.

This chapter quantified key spatial and contextual challenges in Taekwondo annotation through a dataset of 35,913 frames.

From this analysis four key findings were discovered:

1. **Athlete Inactivity:** 68% of frames had containment levels of less than 5%, indicating prolonged inactive phases.
2. **Kick Dynamics**: Median distance decreased by 26% (from 2.3 m to 1.7 m) during kick phases, this also correlates with an increased occlusion risk.

3. **Viewpoint Reliability:** The overhead viewpoint demonstrated balanced dimensions (U/V ratio: 0.96) and minimal occlusion (major containment: 12% of frames), while broadcast viewpoints suffered severe perspective distortion (U/V ratio: 4.16) and frequent occlusion (major containment: 22% in front/back configurations).
4. **Occlusion Severity:** 22% of front/back positional scenarios resulted in major containment (>75%), compared to just 3% in side alignments.

These findings were then used establish a framework to evaluate annotation reliability across three key dimensions:

- **Viewpoint** (overhead, broadcast, audience)
- **Position** (front, side, back)
- **Containment Severity** (none, minimal, partial, moderate, major)

From the analysis conducted in this chapter, three main challenges have been identified in the measurement of interpersonal dynamics in Taekwondo using competition-based footage:

1. High containment presents a significant obstacle, as both manual and automated methods are hindered in scenarios where containment exceeds 75%, which accounts for 22% of front and back interaction cases.
2. Viewpoint bias has been observed, as the aspect ratio (4.16) of broadcast footage is likely to degrade the accuracy of manual and automated approaches as well as introducing greater calibration and transformation error.
3. Dynamic actions such as kicks produce rapid positional changes that lead to occluded scenarios, accuracy of any annotation system may struggle during these events.

Chapter 5 – Reliability of Manual Measurement

5.1 Introduction

In combat sports performance analysis, measurement of athlete positioning and inter-athlete distance is fundamental for tactical evaluation and strategic decision-making. Coaches and analysts routinely use video footage to assess movement patterns, reaction times, and positional strategies, with these measurements forming the basis for performance feedback and training adaptations [94], [95], [96].

Currently, distance and positional measurement in Taekwondo performance analysis relies predominantly on manual annotation by operators. The typical workflow involves an operator calibrating the competition area and manually selecting points representing each athlete's estimated centre of mass projected onto the mat [77]. While this approach is widely used due to its accessibility and low technical requirements, the reliability of these manual measurements under different viewing conditions remains unclear.

Manual annotation introduces inherent subjectivity into performance measurement, as operators must make visual judgments about athlete positioning from 2D video footage [97], [98]. This subjectivity becomes particularly problematic when:

- **Camera viewpoint changes**: Different viewing angles may affect depth perception and measurement reliability
- **Athletes become occluded**: Overlapping or partially hidden athletes complicate position estimation
- **Multiple operators are involved**: Inter-rater variability may introduce inconsistencies in measurement protocols

The extent to which these factors affect measurement reliability has not been systematically quantified in Taekwondo leaving coaches and analysts without clear guidance on when manual annotation is sufficiently accurate for performance analysis purposes.

This chapter aims to quantify the intra and inter- reliability of manual position in competition-based Taekwondo. To provide a more detailed insight, the Framework developed in Chapter 4 is utilised to investigate the impact on viewpoint, orientation and containment on position.

5.2 Method

5.2.1 Experimental Design

This study aimed to assess both intra- and inter-rater reliability of the manual digitisation process described in Chapter 3. A test–retest ratio of 10% of the original dataset was used for this purpose. The reliability assessment followed the same procedures and operational definitions as in Chapter 3, ensuring methodological consistency.

Two reliability conditions were examined:

- **Intra-rater reliability**: The same operator re-annotated the selected dataset after a six-month interval to minimise recall bias.
- **Inter-rater reliability**: A second, independent operator annotated the same dataset using the same instructions and tools.

Due to the initial Intra comparison this left two sets of data for the same 10% of frames. To account for minor systematic differences between rater attempts, inter-rater comparisons were made using the average of both dataset measurements rather than either individual set alone. This approach reduces random error and provides a more stable estimate of true positional agreement.

All procedures were approved by the Sheffield Hallam University Research Ethics Board (ER39637393).

5.2.2 Dataset

The reliability dataset comprised 3,591 frames, representing approximately 10% of the total data collected in Chapter 3. The frames were sampled to reflect a representative spread of movement types, viewpoints, and athlete orientations.

5.2.3 Spatial Calibration and Coordinate System

Digitisation was performed using CVAT. Positions were first expressed in image coordinates (UV) and then mapped to real-world metres via a 2D Direct Linear Transformation (DLT). The eight arena vertices, identified in each video, served as control points to define a consistent transformation from pixel to physical space. This ensured positional consistency across cameras and resolutions.

5.2.4 Annotation Protocol

Athlete position was defined as the projected centre of mass on the mat plane, operationalised as the midpoint between the feet, consistent with Chapter 3. When one foot was occluded, the midpoint was inferred from the visible foot and adjacent frames; when both feet were occluded, interpolation was performed using preceding and following frames and the opponent's location as reference.

Each frame was annotated twice by the same operator (for intra-rater reliability) and once by an independent operator (for inter-rater reliability). All annotations were made independently and without access to prior results.

5.2.5 Statistical Analysis

For each athlete in each frame, positional differences between annotations were computed along both X and Y axes and as a Euclidean distance. Results were summarised using:

- Median difference
- Maximum difference
- 95th percentile difference
- Standard Error of Measurement (SEM)

The SEM represents the standard deviation of measurement error and reflects the typical variability expected if the same measurement were repeated multiple times. A smaller SEM indicates greater measurement precision.

For intra-rater reliability, the SEM quantifies the operator's internal consistency across time. For inter-rater reliability, the SEM indicates the degree to which two independent raters produce comparable results, considering both systematic and random differences.

All analyses followed the scenario framework established in Chapter 4, with results first reported for all frames combined and then stratified by viewpoint and orientation. Inter-rater results were additionally examined by orientation and containment to assess whether reliability varied under different visual conditions.

5.3 Results

5.3.1 Overall Reliability

Table 5.1 presents the overall reliability analysis of all 3,591 annotated frames across intra- and inter-rater comparisons. The table reports the median and maximum absolute differences, the standard error of measurement (SEM), and the 95th percentile for both X- and Y-axis positional differences.

For intra-rater reliability, the median differences were 0.05 m on the X-axis and 0.08 m on the Y-axis. Variability was greater along the Y-axis, where the median difference exceeded the X-axis by 0.03 m and the 95th percentile by 0.13 m (0.39 m versus 0.26 m). The distributions were right-skewed, indicating occasional larger discrepancies, particularly on the vertical axis. The standard error of measurement (SEM) reflected the same trend, with values of 0.08 m on the X-axis and 0.13 m on the Y-axis. Overall, intra-rater uncertainty typically lies within 0.05 - 0.09 m per axis under favourable conditions.

For inter-rater reliability, the median differences increased slightly to 0.06 m on the X-axis and 0.09 m on the Y-axis, representing a rise of approximately 0.01 m on each axis compared to intra-rater results. The Y-axis again showed greater variability, with a median excess of 0.03 m and a 95th percentile difference of 0.12 m (0.46 m versus 0.34 m). SEM values rose correspondingly to 0.11 m and 0.15 m for the X and Y axes, respectively. The inter-rater distributions were also right skewed, with a pronounced tail on the Y-axis where the maximum difference reached 2.15 m, highlighting the influence of occasional outliers.

Intra-rater differences were smaller than inter-rater differences on both axes.

Table 5.1 Intra and Inter-rater reliability of manual position annotation (m, n = 3,591)

Type	Measure	Median	Maximum	SEM	P95
Intra	X-axis	0.05	0.85	0.08	0.26
Intra	Y-axis	0.08	1.37	0.13	0.39
Inter	X-axis	0.06	1.42	0.11	0.34
Inter	Y-axis	0.09	2.15	0.15	0.46

5.3.2 Reliability by Viewpoint

Table 5.2 summarises positional differences by viewpoint. Median absolute differences were smallest for overhead footage, moderate for audience footage and largest for broadcast footage for both intra- and inter- scenarios.

Thus, relative to overhead, the broadcast view increased the Y-axis median by 0.09 m for intra-rater comparisons and by 0.11 m for inter-rater comparisons, whereas audience increased the Y-axis medians by 0.04 - 0.05 m.

On the X-axis the SEM rose from 0.08 - 0.10 m for overhead to 0.10 - 0.14 m for broadcast, with audience at 0.08 - 0.10 m. On the Y-axis the SEM ranged from 0.10 - 0.12 m for overhead to 0.15 - 0.19 m for broadcast, with audience at 0.12 - 0.14 m. The 95th percentile mirrored these differences: for inter-rater comparisons it was 0.30 m on the X-axis and 0.35 m on the Y-axis for overhead, 0.29 m and 0.39 m for audience, and 0.41 m and 0.53 m for broadcast.

P95 values indicated rare but large errors. The largest single-frame Y-axis difference occurred in audience footage for inter-rater comparisons at 2.15 m. The largest X-axis difference occurred in broadcast footage for inter-rater comparisons at 1.42 m. Despite these outliers, overhead consistently produced the smallest medians, SEMs and P95 values on both axes, while broadcast produced the largest values, particularly on the Y-axis.

Table 5.2 Intra- and inter-rater reliability by viewpoint (m)

Viewpoint	Type	Measure	Median	Maximum	SEM	P95
broadcast	intra	X-axis	0.05	0.85	0.10	0.31
broadcast	inter	X-axis	0.07	1.42	0.14	0.41
broadcast	intra	Y-axis	0.13	1.29	0.15	0.44
broadcast	inter	Y-axis	0.16	1.88	0.19	0.53
audience	intra	X-axis	0.05	0.69	0.08	0.24
audience	inter	X-axis	0.06	1.15	0.10	0.29
audience	intra	Y-axis	0.08	1.05	0.12	0.37
audience	inter	Y-axis	0.10	2.15	0.14	0.39
overhead	intra	X-axis	0.03	0.78	0.08	0.17
overhead	inter	X-axis	0.04	1.05	0.10	0.30
overhead	intra	Y-axis	0.04	1.37	0.10	0.27
overhead	inter	Y-axis	0.05	1.94	0.12	0.35

5.3.3 Reliability by Orientation

Table 5.3 summarises positional differences by athlete orientation. Side orientation produced the smallest typical differences on both axes. The intra-rater medians were 0.04 m on the X-axis and 0.07 m on the Y-axis, and the inter-rater medians were 0.05 m and 0.09 m. Front and back orientations showed larger values, with back orientation giving the largest Y-axis medians: 0.09 m for intra-rater and 0.12 m for inter-rater comparisons.

On the Y-axis the SEM for side orientation was 0.12 m for intra-rater and 0.13 m for inter-rater comparisons, whereas front and back orientations were higher at 0.15 - 0.19 m. On the X-axis the SEM for side orientation was 0.07 - 0.10 m, compared with 0.11 - 0.15 m for front and back orientations. For side orientation they were 0.20 - 0.31 m on the X-axis and 0.34 - 0.38 m on the Y-axis. For back orientation they rose to 0.36 - 0.45 m on the X-axis and 0.48 - 0.58 m on the Y-axis. Front orientation showed the highest on the Y-axis, with a 95th percentile of 0.62 m despite a lower median than back orientation.

Table 5.3 Intra- and inter-rater reliability by orientation (m)

Orientation	Type	Measure	Median	Maximum	SEM	P95
front	intra	X-axis	0.05	0.78	0.11	0.32
front	inter	X-axis	0.06	0.90	0.11	0.32
front	intra	Y-axis	0.07	1.37	0.15	0.50
front	inter	Y-axis	0.08	1.94	0.19	0.62
side	intra	X-axis	0.04	0.85	0.07	0.20
side	inter	X-axis	0.05	1.07	0.10	0.31
side	intra	Y-axis	0.07	1.29	0.12	0.34
side	inter	Y-axis	0.09	1.82	0.13	0.38
behind	intra	X-axis	0.06	0.84	0.11	0.36
behind	inter	X-axis	0.07	1.42	0.15	0.45
behind	intra	Y-axis	0.09	1.05	0.15	0.48
behind	inter	Y-axis	0.12	2.15	0.19	0.58

The largest individual errors were observed for the back orientation. The maximum inter-rater difference reached 1.42 m on the X-axis and 2.15 m on the Y-axis. Corresponding maximum for side orientation were 1.07 m and 1.82 m, and for front orientation were 0.90 m and 1.94 m. Across orientations the Y-axis exceeded the X-axis for medians, SEM and P95 values, with the largest median gap between axes occurring for back orientation in the inter-rater comparison at 0.05 m.

These results indicate a clear orientation effect on manual reliability. Side-on views yielded the most consistent annotations, while back orientation produced larger typical and tail errors, particularly on the Y-axis.

5.3.4 Orientation-Containment Reliability

Orientation and containment are measured for every frame; however, certain orientation states are associated with systematically higher containment values, as can be seen in Figure 5.1. Specifically, front and behind alignments typically produce greater bounding box overlap, whereas side alignments tend to produce less, as demonstrated in Chapter 4. The joint analysis therefore examines intra- and inter-rater differences in combined X

and Y displacement, stratified by orientation state and containment level, to assess how agreement varies across these combined conditions.

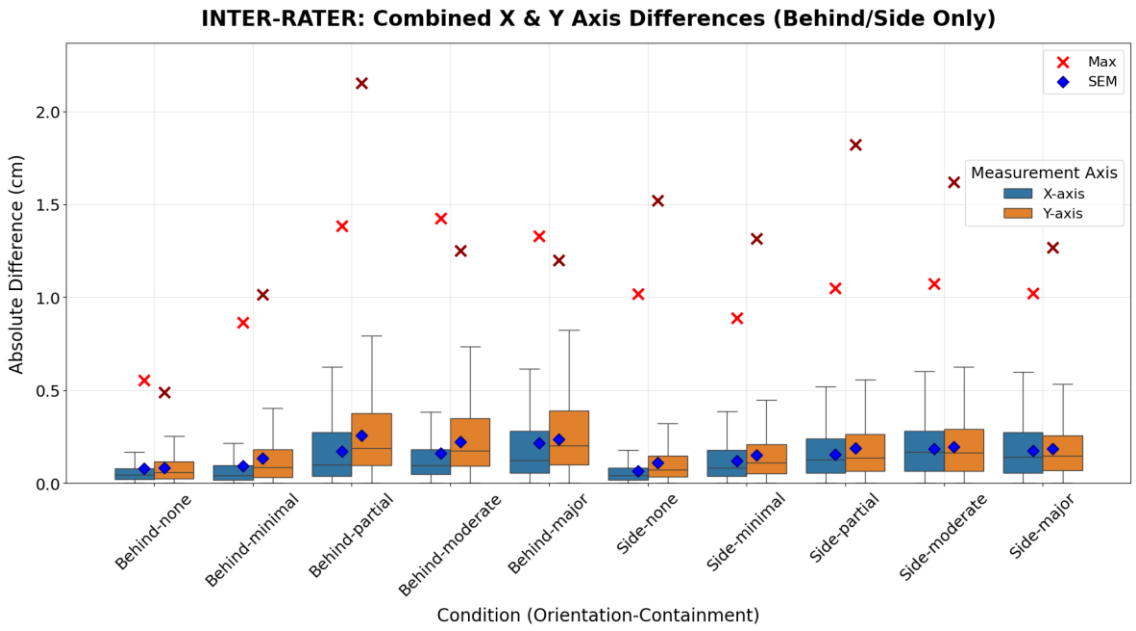


Figure 5.1 The combined X & Y Axis differences stratified by orientation and containment with Max and SEM values.

Inter-rater differences increased with containment severity within both orientations shown. For back orientation, dispersion widened progressively from none through to major containment. The SEM markers rose across the containment bands, and several cells produced maximums that approached or exceeded 2 m on the Y-axis. Side orientation followed the same pattern but at a lower level, with smaller boxes and lower SEMs than the corresponding back-orientation cells. Across all cells the Y-axis distributions were broader than the X-axis distributions, consistent with the axis-wise results in Table 5.3, and the largest extremes were observed on the Y-axis.

At minimal containment, the interquartile ranges were small, and the SEM values were low for both orientations, indicating limited dispersion in the measurements. At partial and moderate containment, there was an increase in spread, reflected by wider

interquartile ranges and higher SEM values. At major containment, the widest distributions and largest maximum errors were observed, particularly for back orientation. These results indicate that the combination of back orientation and higher containment presents the least reliable measurement conditions.

5.4 Discussion

The chapter quantified manual annotation uncertainty for athlete position on the mat plane. Typical axis-wise differences were 0.05 m (X) and 0.08 m (Y) for intra-rater comparisons and 0.06 m (X) and 0.09 m (Y) for inter-rater comparisons, with P95 values of 0.26 to 0.46 m and maximum above 1 m. The disparity between the 95th percentile and the maximum values indicate the presence of occasional large outliers, particularly on the Y-axis, where extreme cases exceeded 2 m despite 95% of frames lying well below 0.5 m. These cases likely reflect momentary occlusions or ambiguity in the definition of foot position under heavy containment, rather than systematic annotation error.

Across all analyses, Y-axis variability exceeded that of the X-axis. This likely arises from camera geometry and parallax effects, which make vertical estimation more sensitive to viewpoint and containment. The difference was consistent across both intra- and inter-rater comparisons, suggesting that this axis-specific uncertainty is inherent to the viewing conditions rather than operator bias.

Camera viewpoint influenced reliability in a manner consistent with camera geometry and occlusion risk. Overhead views reduced perspective effects and lowered occlusion, yielding the smallest medians, SEM and 95th values. Broadcast views increased depth ambiguity along the Y-axis and produced the largest errors. Intra-rater reliability remained systematically higher than inter-rater reliability, confirming that even with clear

protocols, individual perception and interpretation introduce subtle but measurable differences between raters.

Athlete orientation and containment affected agreement in line with visibility of foot contact. Side orientation produced lower error, while front-to-back alignments with partial to major containment increased error, especially on the Y-axis. The joint orientation-containment analysis identified the least reliable conditions were back orientation with moderate or major containment. In these cases, SEM values approached 0.2 m, indicating that repeated measurements under similar conditions could vary by this magnitude even without bias.

Under favourable conditions such as overhead viewpoint with low containment, typical uncertainty was on the order of 0.03 - 0.05 m and 95th percentiles below 0.35 m is achievable with manual annotation. Additionally, under broadcast viewpoints or when back orientation coincides with moderate or major containment, typical Y-axis uncertainty of 0.10 - 0.16 m and 95th percentiles up to 0.5 - 0.6 m are expected.

Limitations should be noted. Positions were estimated in two dimensions using a planar homography, so off-plane movement was not captured. The projected centre of mass was operationalised as the midpoint between the feet, which is a practical proxy rather than a biomechanical estimate. Frame counts differed across orientation-containment cells, which may increase uncertainty in certain stratification groups. Nonetheless, SEM provides a direct quantification of this uncertainty, allowing these conditions to be compared on a standardised scale.

For the research study objective, these results establish quantitative thresholds for evaluating automated systems. Such systems should match or exceed the typical and tail performance observed in manual annotation under comparable conditions and merit closer scrutiny in the challenging combinations identified here. In this way, the present

chapter provides both the empirical foundation and the interpretive framework for assessing automation accuracy in subsequent analyses.

5.5 Conclusion

Manual annotation of athlete position produced small typical differences, with larger errors on the Y axis than on the X axis and greater inter-rater than intra-rater variability. Reliability varied with viewing conditions: overhead views with low containment yielded the highest consistency, while broadcast views and back-oriented athletes under higher containment were least reliable. These findings quantify the limits of manual annotation precision under the tested conditions and summarise the variability that underpins subsequent methodological comparisons.

Chapter 6 – Coach Distance Reliability

6.1 Introduction

Coach-defined distance labels are examined as a practical complement to the manual measurement benchmark. The research studies aim requires automated measurement to be evaluated not only against manual annotation but also against how coaches describe tactical space. Four labels commonly used by coaches are studied: clinch, short, medium and length. The analysis quantifies agreement between coaches and maps each label to measured inter-athlete distances derived from Chapter 3. The objective is to determine whether qualitative labels correspond to stable quantitative thresholds, and whether automation should report numeric distances rather than categories for cross-coach communication**.**

6.2 Method

6.2.1 Data Collection

The data used in this study is a subset from the initial dataset collected in Chapter 3. From the dataset three distinct bouts were selected, each representing the different viewpoints (broadcast, audience and overhead), containing segments with diverse movement patterns, occlusions and distance changes. From this dataset, five short clips of 100 frames (4 seconds) per bout were extracted to capture a range of engagement scenarios, including resets, attacks, and periods of inactivity. These clips represent unique stopping points as shown in Table 6.1. All procedures were approved by Sheffield Hallam University Research Ethics Board (ER39637393).

Table 6.1 – Video clips and their context

Bout	Clip ID	Start Frame	End Frame	Kick	Reset	Note
cb_vs_dp_1	1	375	475	x		Short Kick
cb_vs_dp_1	2	600	700	x		Long Kick
cb_vs_dp_1	3	975	1075		x	Rapid Bearing Change
cb_vs_dp_1	4	1800	1900		x	Standard Reset
cb_vs_dp_1	5	3825	3925			Random Sample
bw_vs_mag_2	6	150	250	x	x	Rapid Distance Change
bw_vs_mag_2	7	750	850		x	Smallest distance recorded (uv)
bw_vs_mag_2	8	1950	2050		x	Largest distance recorded (uv)
bw_vs_mag_2	9	3075	3175		x	Rapid Distance Change
bw_vs_mag_2	10	1350	1450	x		Random Sample
bc_vs_ab_3	11	975	1075		x	Standard Reset
bc_vs_ab_3	12	450	550			Random Sample
bc_vs_ab_3	13	1650	1750	x	x	Rapid Distance Change
bc_vs_ab_3	14	2700	2800			Large Occlusion
bc_vs_ab_3	15	5400	5500	x		Random Sample

6.2.2 Participants

Four elite-level Taekwondo coaches were selected. An elite-level coach in this study was defined as an individual with experience coaching an athlete at an international tournament. Coaches could not discuss the work or their annotations with each other to avoid bias.

6.2.3 Procedure - Video Analysis by Coaches

Coaches reviewed the recorded bouts and classified the distance between athletes using CVAT [32]. The software then displays the video as separate frames which the coach can choose to play continually or select a specific moment. CVAT was configured in tagging mode with the categories: *clinch*, *short*, *medium*, *length*. For each clip, coaches were

asked to: label the starting distance category, play the video and stop it when they believe the categorisation has changed, tag the new category at the transition frame.

Reaction time was not a constraint as the coaches could seek frames freely. This method aimed to simulate how coaches might observe and classify distances during post-competition analysis. Coaches were suggested to watch the video first normally, then re-watch and seek as required.

A sample output of clip ID 1 is shown below in Table 6.2, this is repeated for all coaches for each clip:

Table 6.2 Example coach data

Clip ID	Coach	Frame	Distance
1	A	0	Clinch
1	A	30	Short
1	A	90	Clinch
1	B	0	Short
1	B	25	Length
1	B	80	Short

6.3 Results

From this study the following 'Starting' label was collected. For measuring variability, the same format was used but has not been included for brevity.

Table 6.3: Starting Labels for each clip from four coaches

Bout	Clip ID	Frame	Coach A	Coach B	Coach C	Coach D
cb_vs_dp_1	1	375	Medium	Medium	Short	Length
cb_vs_dp_1	2	600	Length	Medium	Length	Length
cb_vs_dp_1	3	975	Length	Medium	Medium	Length
cb_vs_dp_1	4	1800	Clinch	Clinch	Clinch	Clinch
cb_vs_dp_1	5	3825	Short	Short	Medium	Length
bw_vs_mag_2	6	150	Length	Medium	Length	Length
bw_vs_mag_2	7	750	Clinch	Clinch	Clinch	Clinch
bw_vs_mag_2	8	1350	Length	Medium	Length	Length
bw_vs_mag_2	9	1950	Length	Medium	Length	Length
bw_vs_mag_2	10	3075	Length	Medium	Medium	Length
bc_vs_ab_3	11	450	Length	Length	Length	Length
bc_vs_ab_3	12	975	Length	Length	Length	Length
bc_vs_ab_3	13	1650	Length	Length	Length	Length
bc_vs_ab_3	14	2700	Clinch	Clinch	Clinch	Clinch
bc_vs_ab_3	15	5400	Length	Medium	Length	Length

6.3.1 Inter-Coach Coach Start Label Agreement

To assess the level of consistency among coaches in their initial classification of distance at the beginning of each video clip, agreement was evaluated in two ways: full consensus across four coaches, and pairwise agreement between coaches.

Only 40% of clips showed full agreement among all four coaches, highlighting considerable variability in subjective categorisation even at the start of clips, where the view is typically most stable. This finding shows the challenges of relying on qualitative labels for tactical distance, even among experienced practitioners.

Table 6.4 – Summary statistics of agreement

Total Clips	Full Agreement	Agreement Rate (%)	No Majority	Disagreement Rate (%)
15	6	40%	4	26.6%

Pairwise comparisons as presented in Figure 6.1 show:

- Coaches' agreement ranged from 40% to 87% with the highest agreement seen between coach 1 and coach 4.
- Moderate agreement was observed between coach 3 and both coach 1 and 4 (73%)
- The lowest consistency occurred between coach 2 and coach 4 (40%)

These results suggest that certain coaches may use differing criteria for assigning categories like Length or Medium. The substantial spread in pairwise agreement also supports the idea that personal coaching philosophy or visual interpretation likely plays a role in classification decisions.

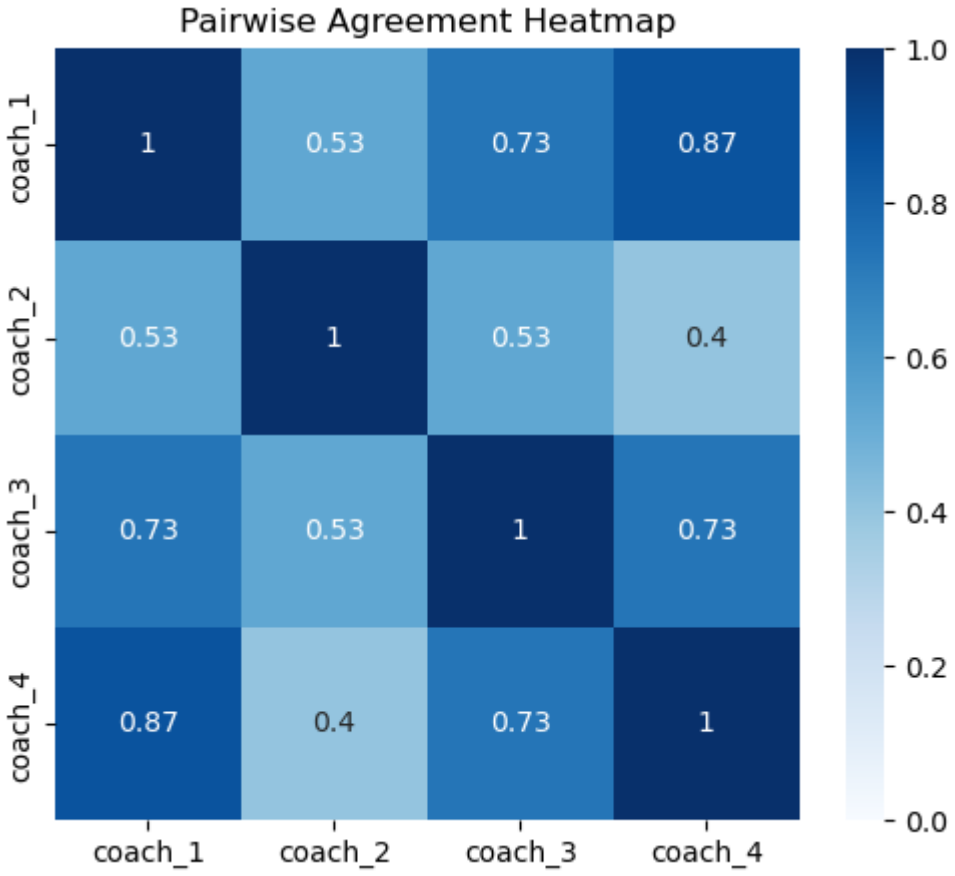


Figure 6.1

The analysis revealed that in 26% of cases no majority consensus where 3 matching coaches could be achieved. These disagreements typically occurred between Medium and Length or between Clinch and Short, where the boundaries are particularly subjective.

A label distribution analysis across all coaches shown in Figure 6.2 revealed a strong preference for the Length classification (55%), followed by Clinch (20%) and Medium (20%). Short was rarely selected (5%), indicating that coaches tend to perceive interactions as either very close, or, clearly spaced with less frequent use of the intermediate zones.

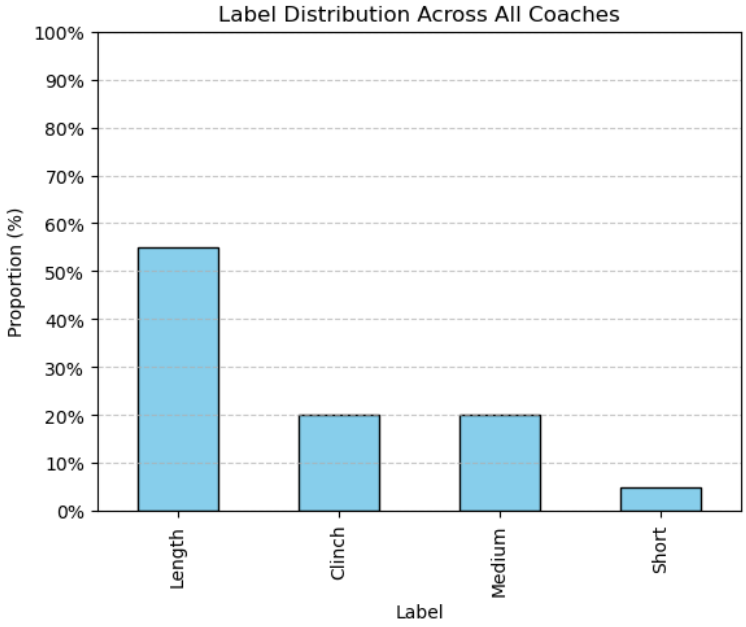


Figure 6.2 – Label Distribution Across All Coaches

6.3.2 Real-World Distance Ranged per Classification

As part of the initial dataset collection in chapter 3, all bouts were digitised for the projected CoM and calibrated using DLT to give spatial measurements in metres. The videos used in this study were drawn from the initial dataset collection. As such, each frame identified by a coach as a transition point could be matched to its corresponding real-world inter-athlete distance. The data for the starting locations is shown in Table 6.3 this is performed for all points where a coach believed the category has changed. In total there are 293 annotations.

Real-world distances were obtained by transforming each frame's projected CoM positions into metres using the transformation matrix established in Chapter 3. For every clip, the frame indices selected by each coach at the start label and at every label change were aligned with the calibrated trajectories. Inter-athlete distance at those frames was computed as the Euclidean separation between athletes on the mat plane. The resulting distributions were summarised per coach and per label.

The distributions show that 1.00 m lies within the interquartile range for Short and within the interquartile range for Medium, while it is above the upper quartile for Clinch. The quartiles for Short were approximately 0.30 - 1.14 m, and for Medium approximately 0.78 - 1.66 m, whereas Clinch had an upper quartile near 0.75 m and a maximum near 0.93 m. Consequently, a 1.00 m engagement was labelled Short by some coaches and Medium by others, with few instances labelled Clinch. This overlap demonstrates that fixed global thresholds are unlikely to match coach practice without coach-specific calibration.

Figure 6.3 shows the distribution of real-world distances associated with each category, broken down by coach. This plot highlights both the central tendencies and variability across coaches as well as the degree of overlap between adjacent categories. This data is also presented in tabular format in Table 6.5.

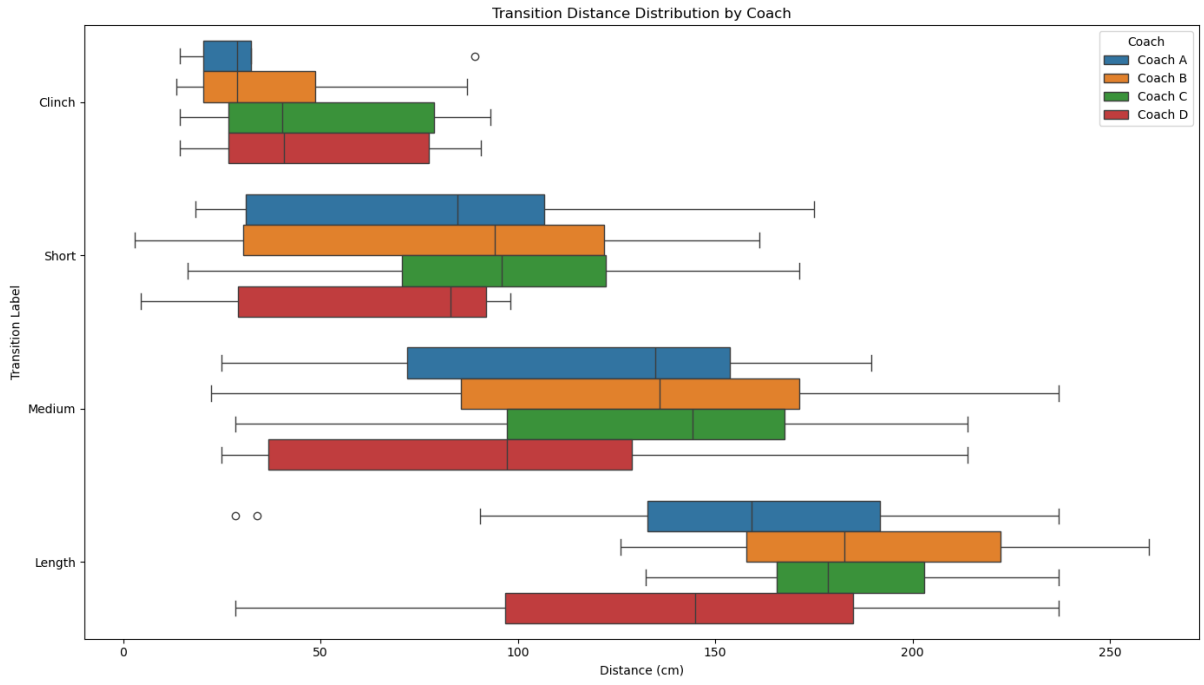


Figure 6.3 – Boxplot of Real-World distances per label, per coach

Table 6.5 Summary statistics of real-world distances by label

	Label	Min	Q1	Median	Q3	Max	Mean	SD
0	clinch	13.6	20.3	31.4	74.9	93.1	44.3	28.5
1	length	28.4	144.8	167.6	202.8	259.7	166.2	50.9
2	medium	22.3	77.8	129.1	165.6	236.9	120.8	57.2
3	short	3.0	30.4	91.7	113.6	175.0	82.8	48.1

Clinch shows the lowest median of 31.4 cm and a mean of 44.3 cm. This category is tightly bound at the lower end but shows a wide interquartile range (IQR), suggesting that some coaches are more conservative about what qualifies as clinch proximity.

Short spans a large range, from as low as 3 cm to as high as 175 cm with overlap into both clinch and medium. This reflects the inconsistency in how short is applied among coaches. This implies there is something additional that some coaches believe that causes a short scenario to be applied, or the potential perspective difference is making it difficult for coaches to consistently agree.

Medium overlaps with both short and length categories, with a wide standard deviation (SD) of 57.2 cm. The IQR ranges from 77.8 cm to 165.6 cm, indicating that this category is also interpreted differently depending on the context or coach.

Length shows the highest upper quartile and maximum distances, as expected. The lower bound of 28.4 cm and the Q1 value of 144.8 cm indicate that there is inconsistency in how this label is applied.

These results reinforce the hypothesis that qualitative distance categories used by coaches do not correspond to universally agreed quantitative threshold. While the categories trend logically in their median tendency (i.e. clinch < short < medium < length), the spread and overlap between the categories highlight that personal interpretation, athlete context and visual perspective all likely play a role in categorising an athlete's distance.

In practical terms this means for automated classification, clear thresholds such as 100 cm is short, 150 cm is medium are unlikely to align with how experts perceive tactical space and in fact coach-specific thresholds should be calibrated to establish these thresholds, or alternatively, coaches are to use a more quantitative measure rather than a judgement based approach when trying to establish distances for training especially in cross-coach communication.

6.4 Discussion

This study evaluated how elite Taekwondo coaches classify interpersonal distances and the reliability of these judgements. The analysis showed moderate agreement at the extremes (clinch and length), but substantial variance in the intermediate zones (short and medium). These findings reflect how subjective and context-dependent coach interpretations can be, even with a consistent tagging framework and the same footage.

The distribution of real-world distances further highlights overlapping use of labels. For distances between 80 and 140 cm were frequently classified as both short and medium, depending on the coach. This ambiguity presents challenges for attempting to standardise and quantify this distance metric both manually and through automated methods while keeping the same coach training approaches.

The consistency in clinch classification suggests that coaches recognise proximity, while length showed a large overlap into other categories, likely due to differences in how coaches interpret disengagements or passive spacing. These patterns point to the need for clearer operational definitions, or calibration methods to map a quantitative metric into that coach's specific thresholds.

6.5 Limitations

Throughout this study several limitations were identified, such as sample size. Only four coaches were involved from the same Taekwondo training programme, which limits generalisability, although shows that even within programme variances can be large. A larger pool of coaches may reveal stronger patterns or more disagreement.

Perspective bias was also a limitation of this study, while coaches all viewed the same clips and as such the same angles per clip, the results were not analysed in terms of perspective, as such differing perspectives such as overhead vs broadcast could influence the perceived spacing.

6.6 Conclusion

The chapter evaluated the reliability of coach-defined distance categories and quantified their correspondence to measured distances. Full agreement among four coaches at clip start occurred in 6 of 15 clips, a rate of 40%. Pairwise agreement ranged from 40% to 87%, indicating wide variability in how distance is categorised. Label usage was dominated by length at 55%, with clinch and medium each at 20% and short at 5%. Real-world distance summaries showed medians of 0.31 m for clinch, 0.92 m for short, 1.29 m for medium and 1.68 m for length, with strong interquartile and range overlap between adjacent labels. At 1.00 m, labels commonly split between short and medium. These results show that qualitative categories do not map to universal quantitative thresholds.

In the context of the research study aim, automated systems should prioritise reporting numeric distances and, where categorical labels are required, should incorporate coach-specific calibration. Chapters 7 and 8 therefore focus on numerical accuracy against the manual benchmark rather than reproducing inconsistent categorical practice.

Chapter 7 – Bounding Box Automation

7.1 Introduction

This chapter evaluates whether a bounding box based method can provide reliable athlete positioning in elite Taekwondo footage when compared with manual annotations of projected centre of mass established as the manual benchmark in Chapter 5. Bounding box detection has been deployed in sport to support automated analysis in cricket, football and tennis, which demonstrates feasibility in high-speed environments, but with sport-specific constraints that limit direct transferability [24], [55], [99]. Off-the-shelf detectors identified people in preliminary tests, yet accuracy deteriorated when Taekwondo specific protective equipment, partial occlusion or non-overhead viewpoints were present. Identity assignment with BoT-SORT and ByteTrack did not resolve mis-association around close contact, which is characteristic of the sport [67], [68].

A custom convolutional detector was therefore trained on Taekwondo competition footage to differentiate the two athletes from other persons near the field of play. Athlete positions were derived as the bottom-centre of each bounding box and expressed in metres to permit direct comparison with manual annotations. Errors were analysed against the Chapter 4 scenario framework, including viewpoint, containment and orientation. The objective is to determine whether bounding box derived positions approach the manual reliability observed in Chapter 5 and to specify the competition conditions under which this method is suitable for applied performance analysis.

7.2 Methodology

7.2.1 Overview

The bounding box based approach provides a practical method for locating athletes within frames by detecting their physical extremities and estimating their CoM using the bottom centre of the bounding box. To validate this method, bounding box derived CoM positions are compared frame-by-frame with manually annotated CoM locations. Errors are calculated across X and Y coordinates in metres. These differences are then stratified according to camera viewpoint, containment and orientation scenarios, as defined in Chapter 4. The performance of the automated system is visualised using Bland-Altman plots to evaluate agreement and identify any systematic biases. All procedures were approved by Sheffield Hallam University Research Ethics Board (ER39637393).

7.2.2 Manual Annotations

The manual annotations used for ground truth comparisons in this chapter originate from the procedures outlined in Chapter 5. Manual operators visually estimated each athlete's CoM as projected onto the mat surface. These annotations were shown to be reliable, with intra-rater and inter-rater median errors ranging between 11 and 13 cm across viewpoints.

This reliability threshold is used as a practical baseline: if the automated method can match or approach this level of accuracy, it may be considered viable for performance analysis.

7.2.3 Bounding Box Generation Process

Bounding boxes were generated through a semi-automated annotation pipeline. First, a pre-trained object detection model was used to automatically detect all persons in each frame of competition footage. These detections were then manually reviewed by manual annotators to ensure accurate bounding box placement. Corrections included:

- Adding missing bounding boxes where detections failed
- Removing false positives (referees, photographers, spectators)
- Adjusting the location and size of bounding boxes for tighter fitting
- Assigning red or blue identities based on athletes' protective equipment

The estimated CoM for each athlete was defined as the bottom centre of their bounding box, corresponding approximately to the athlete's position on the mat. As identity assignment is based on protective equipment colour, separate detection classes for red and blue athletes allow tracking across frames without additional multi-object tracking algorithms. The final output of the bounding boxes can be seen in Figure 7.1.

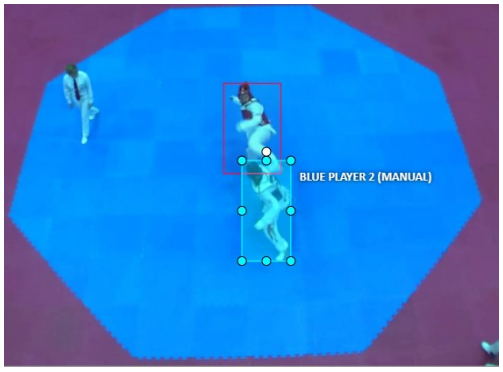


Figure 7.1: Outputs of the annotation process.

7.2.4 Model Training

A custom convolutional neural network (CNN) model was trained specifically for athlete detection in Taekwondo competition footage. Several YOLO architectures (YOLOv7, YOLOv8, YOLOv9, YOLOv10) were benchmarked to determine the most suitable for this application. Model performance was assessed using mean Average Precision (mAP) at Intersection over Union (IoU) thresholds from 0.5 to 0.95, along with precision, recall, and inference speed. The results of these comparisons are presented in Table 7.1. Of the four architectures tested, YOLOv8 achieved the highest mAP of 91.6%. For this reason, YOLOv8 was selected for bounding box detections in subsequent analyses. These model variants were chosen over the alternatives of SSD and FRCNN discovered in the literature as the YOLO architecture performed with a higher mAP when detecting persons in a Taekwondo scene when benchmarked in a pilot study.

7.2.5 Dataset

The dataset comprised 312,493 frames of Taekwondo competition footage, sourced from five major international tournaments. Frames were extracted from 26 matches (78 rounds, 24 athletes; 14 male, 10 female). A total of 186 minutes of video was processed at frame rates of 24, 25, and 30 fps. The dataset was divided into training (70%, 218,745 frames), validation (15%, 46,874 frames), and testing (15%, 46,874 frames). Training data included a wide range of athlete poses, occlusions, and viewpoints, while validation and testing sets were used exclusively for model performance evaluation.

For accuracy assessment, a separate evaluation dataset was assembled (Chapter 3), consisting of 9 rounds of footage with manual CoM annotations, stratified by viewpoint (overhead, broadcast, audience). This ensured independence between training and evaluation data, preventing leakage.

7.2.6 Comparative Analysis

The accuracy of the bounding box model was evaluated by comparing its derived CoM positions against manually annotated positions. For each of the compared frames, positional errors were calculated in metres along the X and Y axis.

These error metrics were then stratified based on the domain framework defined in Chapter 4 by their viewpoint, containment and orientation.

Bland-Altman plots were generated to assess agreement between manual and automated methods, particularly to visualise any systematic bias in over- or under-estimating athlete positions under specific camera conditions.

7.3 Results

7.3.1 Model Training

Several YOLO architectures (YOLOv7, YOLOv8, YOLOv9, YOLOv10) were benchmarked on the Taekwondo dataset. Performance metrics included mean Average Precision (mAP) at Intersection over Union (IoU) thresholds of 0.5 - 0.95, precision, recall, best epoch, inference speed, and training time the results are shown in Table 7.1.

Table 7.1 – Bounding box model training results

Architecture	mAP .5-.95	Precision	Recall	Best Epoch	Inference Speed	Time per Epoch	Training time
YOLOv7	0.891	0.978	0.974	37	8ms	988s	10:09:16
YOLOv8	0.916	0.973	0.969	98	10ms	752s	20:28:16
YOLOv9	0.909	0.959	0.976	17	29ms	2664s	12:45:48
YOLOv10	0.912	0.972	0.912	41	18ms	794	09:02:34

YOLOv8 achieved the highest mAP (0.916) while maintaining high precision (0.973) and recall (0.969). For these reasons, YOLOv8 was selected for bounding box detections in subsequent analyses.

7.3.2 Detection Completeness

Detection completeness was assessed as the proportion of frames in which bounding boxes successfully detected both athletes. Out-of-bounds estimates were treated as failed detections.

Detection completeness as outlined in Table 7.2 varied substantially across conditions. Overhead viewpoint produced the highest detection rate of 99.9%, while broadcast view was lowest at 67.6%. Minimal occlusion maintained a high detection rate of 89.7%, whereas major occlusion reduced completeness to 60.9%. By orientation, front-facing athletes were detected in 94.7% of frames, side orientation in 84.5%, and back orientation

in 71.3%. These results indicate that bounding box detection was reliable, but performance decreased under broadcast viewpoints, severe occlusion, and when athletes were in the behind orientation.

Table 7.2 – Detection completeness by viewpoint, containment and orientation

Group	Category	Frames	Detected Athletes	Detection Rate (%)
viewpoint	broadcast	12865	8703	67.6
viewpoint	audience	12750	11025	86.5
viewpoint	overhead	10263	10248	99.9
containment	none	20781	18461	88.8
containment	minimal	5167	4636	89.7
containment	partial	3666	2836	77.4
containment	moderate	2897	1991	68.7
containment	major	3367	2052	60.9
orientation	front	5126	4853	94.7
orientation	back	6582	4695	71.3
orientation	side	24170	20428	84.5

7.3.3 Overall Positional Accuracy

Bounding box derived centre of mass (CoM) positions was compared with manual annotations. Errors were computed in metres along the X-axis, the Y-axis, and as the resultant distance. As shown in Table 7.3 across all frames, the median resultant error was 0.845 m and the mean resultant error was 1.140 m (SD 0.934 m). Median axis errors were 0.349 m (X) and 0.383 m (Y). Mean axis errors were 0.870 m (X) and 0.436 m (Y), with standard deviations 1.053 m (X) and 0.340 m (Y).

Table 7.3 – Overall positional accuracy of bounding box CoM vs Manual (m)

	Mean	Median	Std
X-axis	0.87	0.35	1.05
Y-axis	0.44	0.38	0.34
Resultant	1.14	0.85	0.93

The mean exceeded the median for all three measures. On the X-axis, the mean error was 0.87 m compared with a median of 0.35 m, and the standard deviation was 1.05 m. This indicates that although most errors were small, there were occasional large X-axis errors that increased the average. In contrast, Y-axis errors were both smaller and more consistent (mean 0.44 m, median 0.38 m, SD 0.34 m). The resultant error followed the same pattern, with a mean of 1.14 m, a median of 0.85 m, and standard deviation of 0.93 m, again reflecting the influence of a minority of larger errors.

7.3.4 Accuracy by Viewpoint

Errors were stratified by camera viewpoint to assess the effect of perspective. Distributions for the X-axis, Y-axis, and resultant distance are presented separately to highlight differences across axes. Table 7.4 shows X-axis error was lowest in overhead and broadcast views (medians 0.11 m and 0.24 m respectively), and much higher in audience footage (median 1.45 m). The interquartile range was narrowest for overhead (0.06 to 0.20 m), intermediate for broadcast (0.10 to 0.41m), and widest for audience (1.09 to 3.07 m). Occasional extreme errors > 5 m occurred in all viewpoints.

Table 7.4 – X-axis error by viewpoint (m)

Viewpoint	Min	Q1	Median	Q3	Max
audience	0.00	1.09	1.45	3.07	5.46
broadcast	0.00	0.10	0.24	0.41	7.40
overhead	0.00	0.06	0.11	0.20	7.71

For the Y-axis as shown in Table 7.5, broadcast produced the largest median error (0.73 m) compared to overhead (0.41 m) and audience (0.18 m). Audience footage had the smallest typical Y-axis errors, though occasional extreme values reached 4.82 m.

Table 7.5 – Y-axis error by viewpoint (m)

Viewpoint	Min	Q1	Median	Q3	Max
audience	0.00	0.08	0.18	0.32	4.82
broadcast	0.00	0.54	0.73	0.92	5.91
overhead	0.00	0.28	0.41	0.53	6.13

Table 7.6 shows that when combining axes into resultant error, overhead produced the lowest median error (0.45 m), followed by broadcast (0.82 m) and audience (1.47 m). Audience footage therefore showed the largest overall error, driven primarily by large X-axis deviations, while broadcast was more affected by Y-axis inaccuracy.

Table 7.6 Resultant error by viewpoint (m)

Viewpoint	Min	Q1	Median	Q3	Max
audience	0.14	1.12	1.47	3.09	6.66
broadcast	0.03	0.63	0.82	0.99	8.44
overhead	0.01	0.34	0.45	0.57	8.63

7.3.5 Accuracy by Containment

Errors were stratified by containment level to assess the impact of occlusion. Five-number summaries for the X-axis, Y-axis, and resultant errors were produced and presented in Table 7.7. X-axis errors were smallest in minimal containment (median 0.23 m) and largest under partial containment (median 0.84 m). Major and moderate containment also produced higher medians (0.50 m and 0.67 m). the no containment group had an intermediate error of 0.31 m. Outliers were observed in all conditions, with some maxima above 7m.

Table 7.7 – X-Axis error by containment (m)

Containment	Min	Q1	Median	Q3	Max
none	0.00	0.10	0.31	1.12	7.71
minimal	0.00	0.10	0.23	1.06	3.95
partial	0.00	0.21	0.84	2.30	3.89
moderate	0.00	0.23	0.67	2.59	3.97
major	0.00	0.21	0.50	1.93	7.40

Table 7.8 shows Y-axis medians were more consistent across containment categories, ranging between 0.31 m (partial) and 0.46 m (major). Moderate and no containment both produced similar medians (0.38 m). Major containment also had the widest spread, with a maximum of 5.91 m.

Table 7.8 – Y-axis error by containment (m)

Containment	Min	Q1	Median	Q3	Max
none	0.00	0.19	0.38	0.61	6.13
minimal	0.00	0.19	0.39	0.58	4.32
partial	0.00	0.15	0.31	0.54	3.93
moderate	0.00	0.21	0.38	0.59	2.39
major	0.00	0.25	0.46	0.71	5.91

Shown in Table 7.9 resultant medians increased with greater containment: 0.70 m under minimal occlusion, 0.82 m under none, 1.05 m under partial, 1.06m under moderate, and 0.99 m under major. Although the ordering is not strictly monotonic, all containment levels with occlusion (partial, moderate, major) showed median errors around or above 1.0 m, compared with < 0.8 m in minimal and no containment conditions. The highest maximum error was observed in the no containment group (8.63 m), reflecting occasional extreme outliers even when no occlusion was present.

Table 7.9 – Resultant error by containment

Containment	Min	Q1	Median	Q3	Max
none	0.01	0.50	0.82	1.21	8.63
minimal	0.02	0.46	0.70	1.18	4.33
partial	0.02	0.60	1.05	2.37	4.13
moderate	0.04	0.58	1.06	2.68	4.18
major	0.01	0.60	0.99	2.08	8.44

7.3.6 Accuracy by Orientation

Bounding box errors were stratified by athlete orientation (front, back, side) to examine whether orientation relative to the camera influenced accuracy.

Shown in Table 7.10 X-axis errors were smallest in back (median 0.25 m) and front orientations (0.26 m), while side orientation produced a larger median error (0.39 m). The interquartile spread was much wider in back orientation (Q3 = 2.84 m) compared to front (1.17 m) and side (1.15 m), suggesting greater variability when athletes were turned away.

Table 7.10 – X-axis error by orientation (m)

Orientation	Min	Q1	Median	Q3	Max
back	0.00	0.09	0.25	2.84	7.03
front	0.00	0.10	0.26	1.17	5.21
side	0.00	0.13	0.39	1.15	7.71

Table 7.11 shows Y-axis medians were consistent across orientations: 0.40 m (back), 0.41 m (front), and 0.37 m (side). Distributions were tighter than for the X-axis, with upper quartiles below 0.63 m in all cases.

Table 7.11 – Y-axis error by orientation (m)

Orientation	Min	Q1	Median	Q3	Max
back	0.00	0.23	0.40	0.57	6.13
front	0.00	0.25	0.41	0.57	5.09
side	0.00	0.17	0.37	0.63	5.91

Table 7.12 shows resultant medians were lowest for front orientation (0.70 m), close to back (0.72 m), and highest for side (0.88 m). Back orientation also produced the widest spread, with an interquartile range extending up to 2.88 m and a maximum error of 8.63 m.

Table 7.12 – Resultant error by orientation (m)

Orientation	Min	Q1	Median	Q3	Max
back	0.01	0.48	0.72	2.88	8.63
front	0.02	0.47	0.70	1.27	7.28
side	0.02	0.53	0.88	1.25	8.44

Orientation influenced error magnitude less than viewpoint or containment, but side orientation consistently yielded larger errors, and back orientation produced greater variability.

7.3.7 Error Distributions

To illustrate the spread of positional errors, histograms of resultant error were produced for all frames combined and separately by viewpoint.

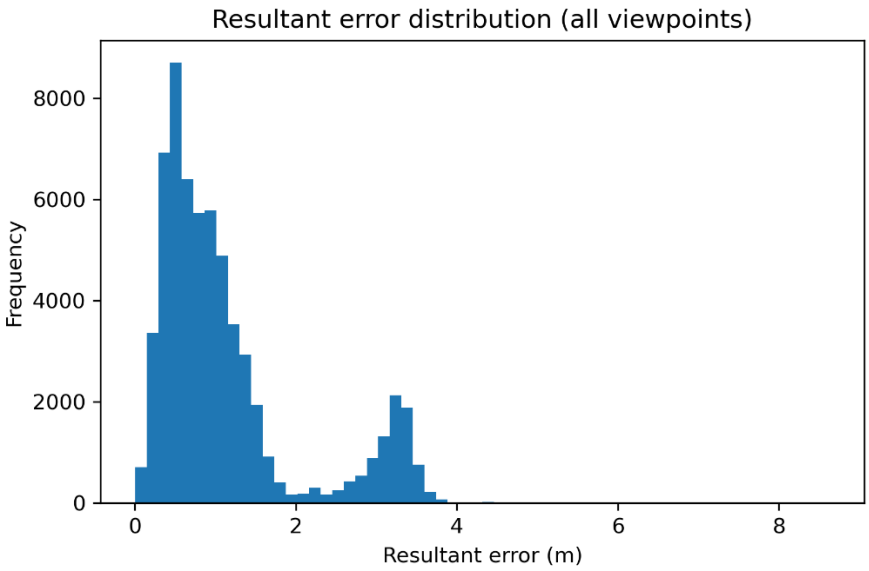


Figure 7.1 Histogram of resultant error across all viewpoints

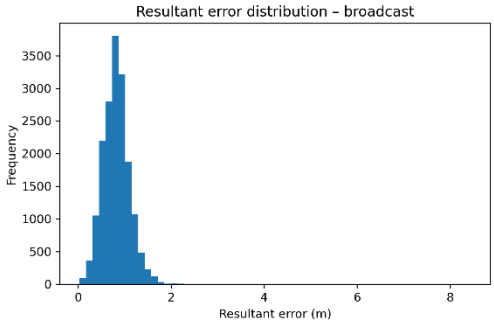


Figure 7.2a - Broadcast

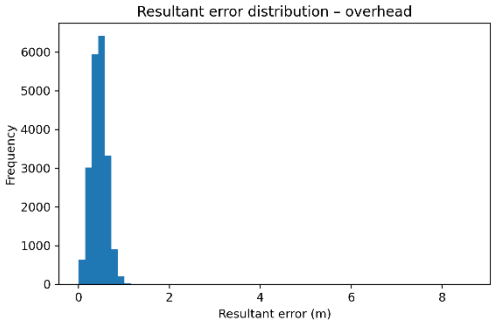


Figure 7.2b - Overhead

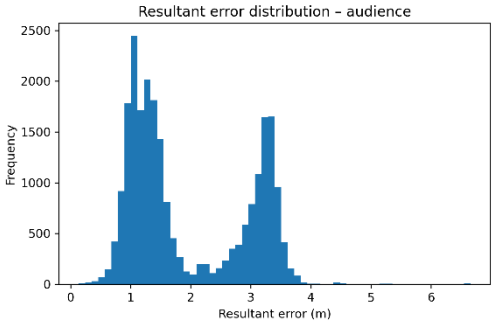


Figure 7.2c Audience

Across all frames (Figure 7.1), most errors were between 0.5 m and 2.0 m. The distribution was positively skewed, with a small number of outliers above 4 m.

When stratified by viewpoint (Figure 7.2), broadcast errors (panel a) were unimodal and concentrated between 0.2 m and 1.5 m, with most frames clustering around 0.8 to 1.0 m. Overhead errors (panel b) were tightly distributed below 1.0 m, with the highest density between 0.2 m and 0.6 m, reflecting the consistency of this viewpoint. In contrast, audience errors (panel c) displayed a bimodal distribution with peaks around 1.2 m and 3.2 m, indicating that while many frames were accurate within 1-2 m, a distinct group of cases produced much larger errors.

Together, the data shows that overhead and broadcast viewpoints produced compact and unimodal error distributions, while audience footage exhibited greater variability and a second mode of larger errors.

7.3.8 Bland–Altman Analysis

Agreement between bounding box derived and manually annotated positions was assessed using Bland–Altman analysis, which visualises the difference between two measurement methods against their mean. Bias and 95% limits of agreement (LoA) are reported for the full dataset and for a representative challenging condition (broadcast with partial occlusion). The straight-line outlier shown in Figure 7.3 is a clipping error whereby the athlete had walked outside of the calibrated area in a paused state by their coaches.

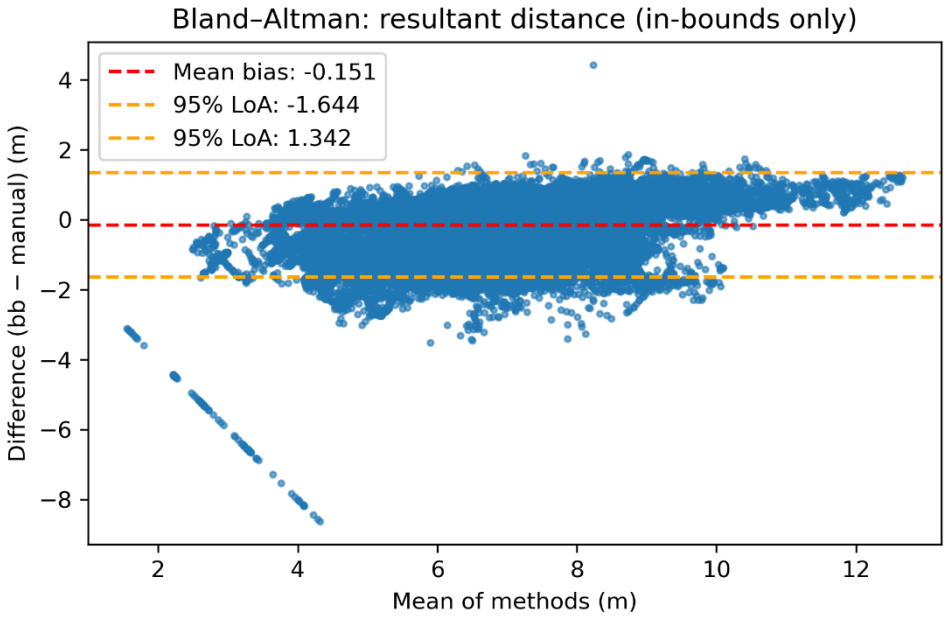


Figure 7.3 – Bland–Altman plot of resultant distance (all frames, in-bounds only)

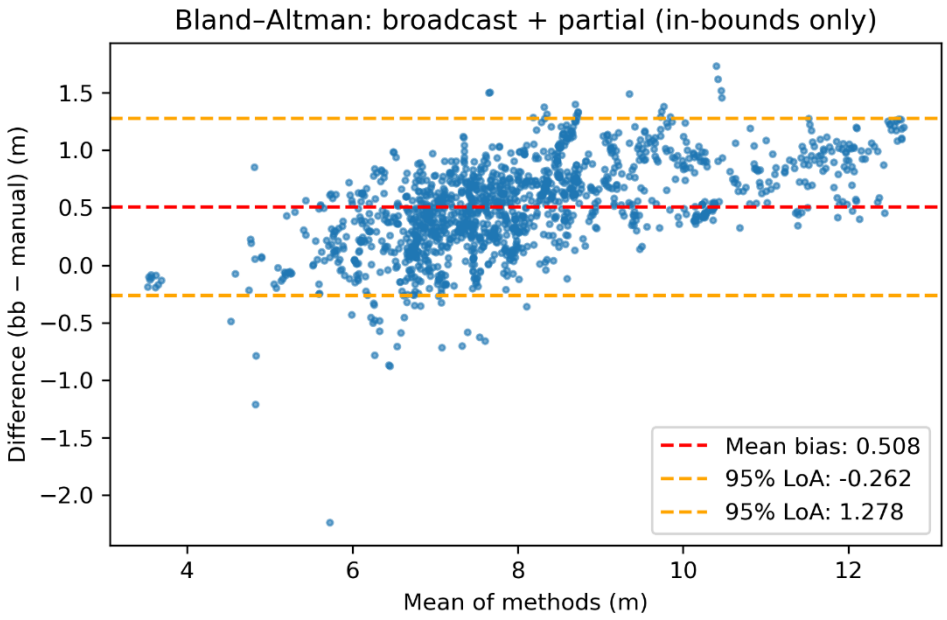


Figure 7.4 – Bland–Altman plot of resultant distance (broadcast with partial occlusion, in-bounds only)

Across all frames (Figure 7.3), the mean bias was -0.151 m, with 95% limits of agreement from -1.64 m to 1.34 m. This indicates that on average bounding box estimates underestimated the manual CoM by approximately 0.15 m, with most differences lying within ±1.5 m.

For the broadcast with partial occlusion condition (Figure 7.4), the mean bias increased to 0.50 m, with narrower 95% limits of agreement from -0.26 m to 1.27 m. This shows that in this condition bounding boxes systematically overestimated CoM position by approximately 0.5 m, while the range of variation remained slightly over ±1 m.

Taken together, the Bland-Altman analyses demonstrate that bounding box estimates were close to unbiased overall, but agreement intervals extended beyond ±1 m. In more challenging conditions such as broadcast with partial occlusion, systematic error increased, and variation remained substantial.

7.3.9 Investigation of larger errors

A qualitative investigation was conducted to understand the causes of large errors (> 2.5m) in the audience viewpoint. Within the viewpoint 5,544 (43.5%) of frames displayed resultant errors exceeding 2.5m. Analysis of representative high-error cases revealed three primary failure modes: perspective distortion, stability and misdetection.

A significant proportion of large errors occurred when athletes were positioned near the far edge of the mat with reference to the camera. Analysis revealed pixel-to-meter ratios ranging from 10 to 131 px/m along the Y-axis. Athletes positioned in the far from the audience camera showed significantly higher errors whereby physically less pixels were representing a larger distance. Figure 7.5 shows a representative case where the athlete positioned at the edge of the plane experiences a 2.7m error while the difference in UV space is minimal.

Figure 7.5 – Edge of Plane

Periods of these errors experienced minor camera stability issues during the 2-minute bout lasting around 5 seconds. Upon visual inspection of the footage revealed camera movement caused by audience members walking past the camera position. This degraded the calibration accuracy, as the camera position no longer corresponded to the pre-recorded calibration parameters. While it is difficult to represent movement in a still frame, figure 7.6 shows a representative case with 3.8m error, while the athletes and predicted positions were visually OK. It is possible to partially see the audience member at the bottom of the frame.

Figure 7.6 – Audience degradation

Occlusion-related errors accounted for the remainer of large errors. These occurred when the target athlete was partially or fully occluded by their opponent, and another athlete of the same team colour was visible in the background. The detection system occasionally misidentified the background athlete as the target, resulting in substantial position errors.

Figure 7.7 demonstrates a representative mis-detection case where the blue athlete in the foreground is occluded by their opponent, while another blue athlete walks in the

background. The detection system incorrectly assigned the bounding box to the background athlete, resulting in a large positional error. It would be possible to implement a post CNN filter these coordinates out using geo-fencing, however this is the true raw output from a CNN based system.

Figure 7.7 – Mis-detection error

7.4 Conclusion

Bounding box derived positions were compared with manual annotations to judge fitness for purpose against the research study aim. Agreement analysis indicated an overall bias of -0.15 m with 95% from -1.64 m to 1.34 m, and a positive bias of 0.50 m with limits from -0.26 m to 1.27 m in broadcast footage with partial occlusion.

Relative to the manual benchmark reported in Chapter 5, the method did not achieve typical manual axis-wise errors. It is concluded that bounding boxes are suitable for coarse positional trends where overhead footage is available and near-complete coverage is required, and unsuitable for precise positional quantification in broadcast or audience viewpoints, particularly when partial or full containment is present. The next chapter evaluates whether pose estimation reduces error towards the manual benchmark while retaining adequate frame coverage.

This chapter therefore addresses research objective 5 by assessing the positional accuracy and practical viability of CNN-based bounding box models, including their limitations across viewpoints and containment conditions. In doing so, it also partially fulfils research question 3 by establishing the levels of positional accuracy achievable with a bounding box approach and identifying the conditions under which the method is viable for applied analysis. This addresses the research study aim by showing that bounding boxes alone do not meet the reliability requirement for precise positional measurement except in overhead footage.

Chapter 8 – Pose Estimation Automation

8.1 Introduction

A pose-estimation approach is evaluated to determine whether accuracy can be improved over the bounding box method while maintaining applicability to real competition footage. Athlete identity is constrained using the detector from Chapter 7; pose keypoints are then used to compute projected centre of mass in calibrated metres. Errors are analysed by viewpoint, occlusion and orientation to align with the Chapter 4 scenario framework. The objective is to establish whether pose estimation reaches or approaches the manual benchmark defined in Chapter 5 and to specify the conditions under which it should be preferred to bounding box only estimates.

8.2 Methodology

8.2.1 Overview

This study evaluates the OpenPose BODY_25 model for positional estimation of Taekwondo athletes during competitive bouts. Athlete identity is assigned using bounding box detections from the model developed in Chapter 7. CoM positions are derived from pose keypoints and compared to manual annotations to quantify positional error. Error metrics are stratified by viewpoint, containment, and orientation as defined in Chapter 4.

8.2.2 Data Collection

The dataset comprises competition footage previously used in Chapter 7, sourced from major international Taekwondo tournaments including the European Championships and the World Taekwondo Grand Prix. All footage contains elite-level athletes, with multiple weight categories and both male and female competitors. Frames were sampled at every 10th frame from each bout to ensure coverage of diverse engagement scenarios while maintaining manageable data volume. Viewpoints represented are broadcast, audience and overhead. Each sampled frame was previously classified for athlete orientation and containment level according to the framework in Chapter 4. All procedures were approved by Sheffield Hallam University Research Ethics Board (ER39637393).

8.2.3 OpenPose Model Configuration

Pose estimation was performed using the OpenPose BODY_25 model (version 1.7.0). The network input resolution was set to 656 × 368 pixels with a confidence threshold of 0.5 applied to remove low confidence keypoint predictions. Original frame resolution was preserved during inference; no cropping, resizing, or normalisation was applied. Model outputs were exported in JSON format and subsequently converted into CSV files for analysis. For each person detected, the BODY_25 keypoints was recorded along with their associated confidence values.

8.2.4 Athlete Identification

Athlete identity was assigned by matching pose keypoints to bounding boxes generated in Chapter 7. For each detected pose, the number of keypoints located within a given bounding box was counted. The pose with the highest count within each bounding box was assigned that athlete's identity ("red" or "blue"). Poses with fewer than 50% of keypoints inside any bounding box was excluded. Any poses not associated with either athlete was disregarded.

8.2.5 Centre of Mass Calculation

The CoM for each athlete was estimated from pose keypoints. The calculation followed the definition used for manual annotations in Chapter 5: the midpoint between the left and right ankle keypoints was projected onto the mat plane and calibrated to real-world coordinates. This ensured direct comparability between manual and automated measurements.

8.2.6 Spatial Calibration

Pixel coordinates were transformed to real-world coordinates using the Direct Linear Transformation (DLT) method described in Chapter 3. This utilised known reference points on the competition area, allowing CoM positions to be expressed in metres relative to the mat.

8.2.7 Accuracy Assessment

Positional error was calculated for each athlete in each frame by subtracting the manual CoM coordinates from the OpenPose derived coordinates. Errors was computed along the X-axis, Y-axis, and as a resultant distance. Detection completeness was calculated as the percentage of frames in which both athletes was successfully assigned a valid CoM. Results were stratified by the framework defined in Chapter 4.

For each condition, mean, median, and standard deviation of errors was reported. Agreement between manual and automated methods was examined using Bland–Altman analysis to identify any systematic bias.

8.3 Results

8.3.1 Detection Completeness

Table 8.1 shows our detection rate for variable viewpoints. Detection rate varied notably with viewpoint, containment level, and orientation. The highest detection rates occurred in frames without occlusion, whereas full occlusion reduced the proportion of frames with valid CoM for both athletes to around one-third. Audience viewpoint produced the highest completeness, while overhead had the lowest. Side-facing athletes was detected more consistently than those front or back orientations.

Table 8.1 - Detection completeness by viewpoint, containment, and orientation

Condition	Frames	Both Athletes Detected	Detection Rate (%)
Broadcast	1291	1109	85.9
Audience	1275	1139	89.3
Overhead	1029	825	80.2
No Occlusion	2101	2047	97.4
Partial Occlusion	509	440	86.4
Full Occlusion	339	119	35.1
Side	2415	2242	92.8
Front	517	397	76.8
Behind	663	434	65.5

These results suggest that OpenPose is most reliable when occlusion is absent and the athletes are side-on to the camera, with audience-view footage providing the most consistent detection.

8.3.2 Overall Positional Accuracy

Table 8.2 shows across all frames, positional accuracy was higher for the X-axis than for the Y-axis, with median errors of 0.09 m and 0.32 m respectively. The mean resultant error was over half a metre, indicating that while many detections aligned closely with manual annotations, a subset of frames exhibited large deviations.

Table 8.2 - Overall positional accuracy of pose CoM estimates compared to manual.

Axis/Metric	Mean Error (m)	Median Error (m)	Std Dev (m)
X-axis	0.18	0.09	0.31
Y-axis	0.45	0.32	0.49
Resultant	0.52	0.39	0.56

The higher Y-axis error suggests greater challenges in estimating depth compared to lateral positioning, particularly in broadcast and audience views where perspective distortion is more pronounced.

8.3.3 Accuracy by Viewpoint

Table 8.3 shows viewpoint strongly influenced performance. Overhead footage produced the lowest mean and median errors across all axes, while broadcast footage performed worst, primarily due to large Y-axis errors. Audience view offered intermediate accuracy.

Table 8.3 - Positional accuracy by camera viewpoint

Viewpoint	Mean X (m)	Median X (m)	Mean Y (m)	Median Y (m)	Mean Resultant (m)	Median Resultant (m)
Broadcast	0.22	0.06	0.74	0.62	0.83	0.68
Audience	0.20	0.14	0.41	0.30	0.47	0.35
Overhead	0.11	0.05	0.17	0.11	0.22	0.14

This pattern is consistent with expectations, as overhead footage minimises perspective effects and reduces occlusion from other participants.

8.3.4 Accuracy by Containment

Table 8.4 shows the extent of occlusion had a clear impact on positional accuracy. Mean resultant error increased progressively from no occlusion to full occlusion, more than doubling in the latter case. Y-axis estimates were consistently more affected than X-axis estimates.

Table 8.4 – Accuracy by Containment

Containment	Mean X (m)	Median X (m)	Mean Y (m)	Median Y (m)	Mean Resultant (m)	Median Resultant (m)
None	0.15	0.07	0.39	0.31	0.45	0.37
Partial	0.20	0.10	0.49	0.32	0.56	0.36
Full	0.52	0.45	0.89	0.50	1.13	0.85

8.3.5 Accuracy by Orientation

Table 8.5 shows orientation effects were less pronounced but still present. Front oriented athletes yielded the lowest mean resultant error, while side orientation resulted in the highest, driven by increased Y-axis inaccuracies. In the X-axis Side had a lower error than behind.

Table 8.5 – Accuracy by orientation

Orientation	Mean X (m)	Median X (m)	Mean Y (m)	Median Y (m)	Mean Resultant (m)	Median Resultant (m)
Side	0.18	0.08	0.49	0.36	0.55	0.43
Front	0.14	0.08	0.29	0.23	0.35	0.27
Behind	0.23	0.15	0.37	0.29	0.46	0.35

8.3.6 Bland–Altman Analysis

Bland–Altman analysis was conducted to assess agreement between OpenPose derived and manually annotated resultant CoM positions. Across all frames, the bias was small, but the limits of agreement indicated potential for substantial errors in individual frames.

In the broadcast with partial occlusion condition, both bias magnitude and limits widened, indicating reduced agreement in more challenging conditions.

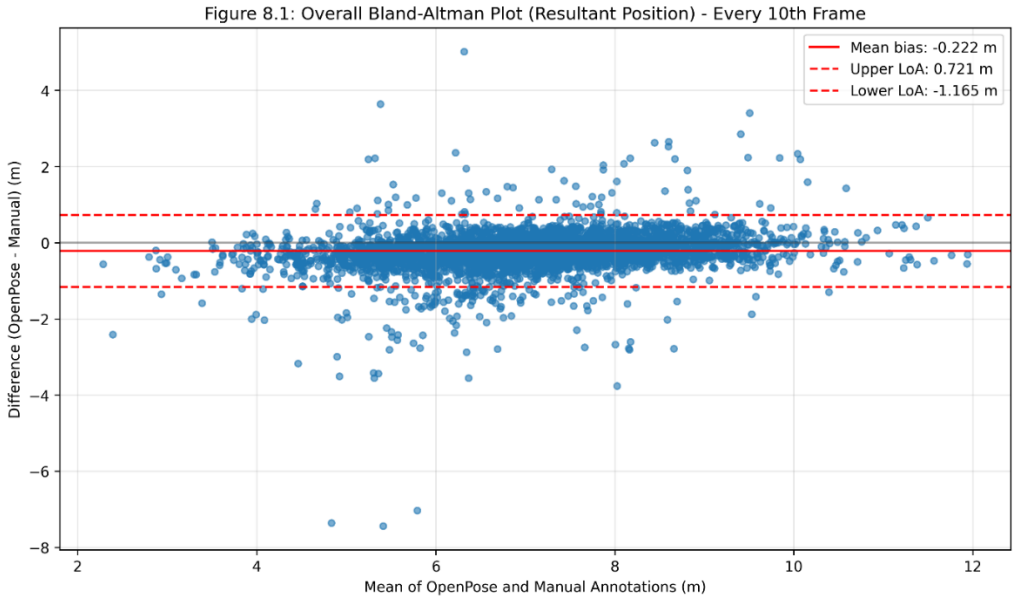


Figure 8.1 overall bland-Altman plot (resultant position) – every 10th frame

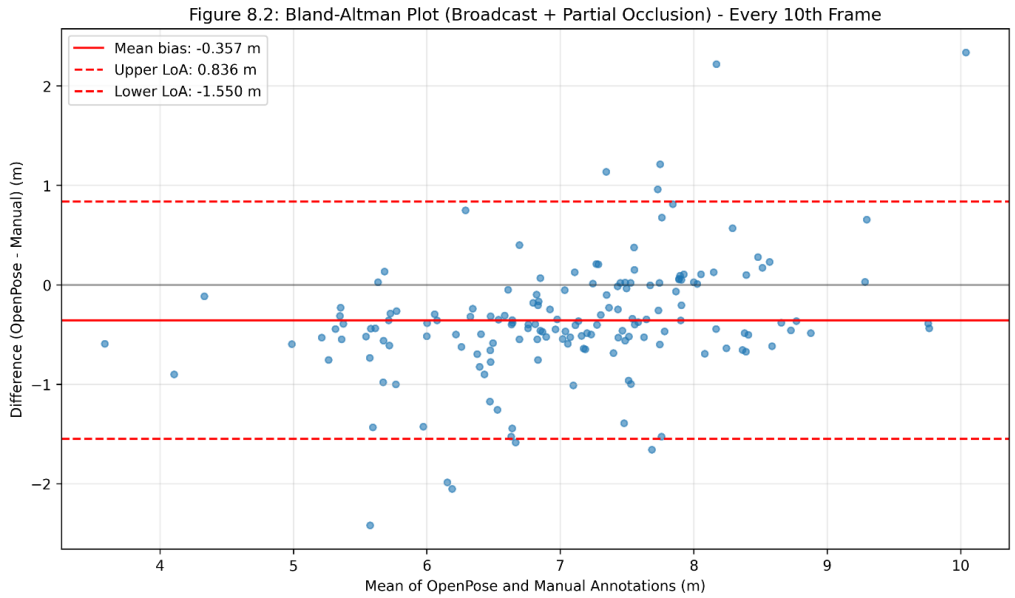


Figure 8.2 bland-Altman plot (broadcast + partial occlusion) - every 10th frame

8.4 Discussion

This chapter evaluated OpenPose, used in combination with bounding box based identity assignment, for estimating the centre of mass (CoM) of athletes in elite Taekwondo competition footage. Performance was assessed in terms of detection completion, positional accuracy under different conditions including viewpoint, orientation and containment, and agreement with manual annotations.

Detection completeness was highest in frames without occlusion at 97.4%. In side-on orientation dropped to 92.8%. Full occlusion reduced this figure to only 35.1%. Behind orientation also proved challenging, with a completeness rate of 65.5%. Among viewpoints, the audience angle produced the most consistent detections at 89.3%, while overhead achieved the lowest with 80.2%. Despite its advantage in positional accuracy, overhead may lose completeness because athletes occupy a smaller portion of the frame at the resolution used here.

Overall, the median resultant error was 0.39 m, and the mean was slightly higher at 0.52 m. Errors was smaller along the X-axis (lateral positioning) than the Y-axis (depth), which is consistent with the increased influence of perspective distortion on depth measurements. This pattern held across all conditions tested.

8.4.1 Viewpoint

Viewpoint was a major factor in determining accuracy. Overhead footage produced a mean resultant error of 0.22 m, lower than audience (0.47 m) and broadcast (0.83 m) views. The high broadcast error was driven by Y-axis inaccuracies, reflecting the inherent difficulty of estimating depth from a shallow viewing angle.

8.4.2 Containment

When no occlusion occurred, mean resultant error was 0.45 m; partial occlusion raised this to 0.56 m, and full occlusion more than doubled it to 1.13 m. This pattern shows that positional reliability deteriorates once joint visibility is reduced. Y-axis estimates suffered most, suggesting that obscured lower limbs or body overlap cause depth estimation to perform worse.

8.4.3 Orientation

Front oriented athletes were the most accurately measured with 0.35 m mean resultant error. Side orientation produced the largest errors at 0.56 m, which may be due to misplacement of ankle keypoints during lateral kicking actions. Behind orientation was intermediate at 0.46 m, reflecting both reduced visibility and fewer keypoints available for CoM calculation.

8.4.4 Agreement

Bland–Altman analysis confirmed these patterns. Across all frames, the mean bias was –0.22 m, with wide limits of agreement (–1.17 m to 0.72 m). In practical terms, most measurements were close to manual annotations, but instances of over 1 m were observed. In the broadcast with partial occlusion condition, bias increased to –0.35 m, and the limits widened (–1.55 m to 0.84 m). This demonstrates the compounding effect of an unfavourable viewpoint and visual obstruction.

Large errors were concentrated when depth had to be inferred from broadcast geometry and during full containment, which produced a median resultant error of 0.85 m and wide Bland–Altman limits. This pattern indicates that the principal limitation is depth reconstruction from monocular views rather than lateral localisation.

Pose estimation delivers the best results in overhead, unobstructed conditions. It struggles in broadcast views, especially when athletes overlap or when joints are hidden. A selective, hybrid approach whereby bounding boxes ensure robust detection and pose estimation refines position only when the view is favourable may offer the most reliable tracking in Taekwondo competition analysis.

8.5 Conclusion

The chapter evaluated pose-based position estimates with detector-assisted identity assignment. Across all frames, the median errors were 0.09 m on the X-axis, 0.32 m on the Y-axis and 0.39 m for the resultant distance. Viewpoint effects were pronounced: median resultant error was 0.14 m for overhead, 0.35 m for audience and 0.68 m for broadcast. Occlusion degraded accuracy from a median of 0.37 m without occlusion to 0.85 m under full containment. Agreement analysis indicated a mean bias of approximately −0.22 m, with limits of agreement spanning roughly −1.17 m to 0.72 m.

Relative to the manual benchmark, pose estimation matched or exceeded typical manual axis-wise medians under overhead and audience viewpoints without heavy occlusion, but performance deteriorated under broadcast and full occlusion. Consequently, pose estimation should be used for fine positional analysis when viewpoint and visibility are favourable, and avoided for precise distance estimation in broadcast footage with substantial occlusion.

Chapter 9 – Discussion

9.1 Manual Annotation Reliability and Benchmarking

The study began by establishing a manual performance benchmark through manual video annotation of athlete positions. The results demonstrated that when conditions were ideal: clear views, minimal occlusion, and clear task definitions, operators could achieve high intra- and inter-rater consistency. This aligns with the sports analysis literature, which has long used manual notational analysis as a gold standard despite its labour intensity. Prior work indicates that video analysts can be reliable in unambiguous scenarios; for example, time-motion studies in combat sports often report almost perfect agreement on basic event counts or phase durations when visual cues are clear [100]. Frames with minimal containment were annotated consistently across raters, echoing the notion that operators excel when the subjects are fully visible and distinguishable.

The findings also highlight the contextual limits of manual annotation. Reliability decreased in challenging scenarios for example during close engagements when one athlete partially or fully occluded the other, or when camera angles were at a low angle. Annotators sometimes disagreed on the exact athlete positions in these instances, reflecting how perception can struggle with visual ambiguity. This is analogous with other studies where expert tennis coaches showed strong agreement when assessing clearly observable aspects of technique but had more difficulty reaching consensus on subtle or occluded movements (e.g. shoulder rotation speed) [93]. Likewise, a recent reliability study of kickboxing match analysis reported ICC values of 0.88 – 0.99 for straightforward metrics like counts of punches/kicks, but lower agreement (ICC down to 0.16) for subjective classifications of strikes [36]. These parallels suggest that Taekwondo annotators' inconsistencies under difficult conditions are not an anomaly but rather part

of a general pattern: manual reliability is context dependent and degrades when the visual scene is complex. The scenario-stratified analysis (Chapter 5) confirmed this, near perfect agreement in easy scenarios was but notably reduced consistency in high-occlusion frames and distorted camera views. Such results reinforce the literature's warnings that manual annotation, while accurate in principle, is prone to error in noisy real-world footage.

This manual reliability analysis represents a novel contribution to knowledge within Taekwondo performance analysis. No previous research in the sport has established a comprehensive, scenario-based manual benchmark for positional measurement, nor quantified how reliability varies across viewpoints, occlusion levels, athlete alignments and kicking phases. Existing literature reports only a single aggregated ICC value for athlete position tracking, without any stratification by the visual or tactical contexts that characterise real competition footage. By contrast, the present study provides the first detailed breakdown of manual reliability under authentic, elite-level conditions, establishing a multi-dimensional baseline against which automated systems can be validly compared. This contribution is critical, because it defines the upper bound of human measurement performance in Taekwondo and creates the first domain-specific reference standard for evaluating CNN-based approaches.

9.2 Coach Distance Classification

The consistency of elite coaches in classifying the interpersonal distance using their own qualitative categories (e.g. clinch, short, medium, length) was analysed. This was motivated by the widespread use of such categorical distance schemas in Taekwondo coaching, despite a lack of standardisation in the literature. Coaches often develop intuitive thresholds for what they consider "critical distance" in sparring, but until now these thresholds had not been rigorously mapped to actual metric values or tested for

inter-coach agreement. Findings revealed a moderate level of consistency: coaches broadly agreed on the rank order of distances (for example what situations count as clinch vs length), but there was noticeable variation in the exact boundary definitions each coach used for these categories. In practical terms, one coach's "medium distance" could overlap with another coach's "short distance" in absolute units, underscoring the subjective nature of these labels. This result is consistent with other literature which noted that different Taekwondo analysts often use different scales or criteria in match analysis, leading to inconsistent data and interpretations [77]. This study addresses this gap by establishing numeric ranges for each coach-defined category, effectively creating a translation between the coaches' language and measured distances.

9.3 CNN-Based Bounding Box Tracking Performance

Chapter 7 focused on an automated bounding box method. A custom object detection model (based on YOLO architecture) was trained to detect and localise the two athletes in each frame, using discrete categories for the red and blue protector to maintain each athlete's identity over time. In optimal conditions (both athletes fully in frame, minimal overlap), the bounding box tracker achieved high accuracy in estimating the distance between fighters, often within a few centimetres of the manually annotated positions. This is in line with other sports tracking systems in other domains: for example, the FootyVision project in football achieved extremely high detection precision (mAP ~95.7%) and excellent multi-object tracking accuracy (MOTA ~94%) by using an advanced one-stage CNN detector (YOLOv7) on football footage. Effectively extending the success of object detection from sports like football and basketball into the realm of martial arts, confirming that bounding box-based tracking can generate useful positional data even during the fast, dynamic movements of sparring.

The analysis also showed the failure modes of the bounding box approach, most of which stem from the same challenges that operators faced. One major issue was occlusion: during close clinches or momentary overlaps (e.g. one athlete executing a turning kick and momentarily occluding the other), the detector would sometimes merge the two fighters into one box or temporarily lose one athlete. This led to discontinuities in the distance measurement or, in worst cases, identity switches when the occluded athlete reappeared. The sensitivity to occlusion is a well-documented limitation of vision trackers [101]. Prior systems often resort to multi-object tracking (MOT) algorithms to re-identify targets across frames, but as the literature review noted, standard MOT can struggle when two targets interact closely or swap positions. In line with recommendations by other researchers, by integrating domain-specific cues to perform identity tracking by leveraging the distinct colours of the athletes' protective gear (red vs. blue) as an additional feature for ID consistency [102], [103]. A similar approach was hinted at in an MMA study suggested using glove/short colours to keep track of fighters in grappling exchanges [64]. The results in this study showed that such custom solutions improved the robustness of the bounding box tracker, reducing instances of identity swaps compared to a generic MOT method. Overall, the bounding box method proved reliable in many scenarios and served as a solid backbone for the tracking system, but its performance degrades in exactly those hard cases that one would expect heavy occlusion, identical appearances, and extreme camera angles. These findings reinforce the literature's consensus that no single method will flawlessly handle all cases in unconstrained sports video.

9.4 Pose Estimation Model Performance

Complementing the bounding box approach, CNN-based pose estimation model (OpenPose) was evaluated a to track athletes' keypoints and derive their positions. Pure pose estimation offers more fine-grained analysis, not just where an athlete is, but how their body is oriented and moving. In terms of measuring the distance between athletes (the primary metric), the pose approach tracked each athlete's centre of mass (CoM) via skeleton keypoints and then computed inter-CoM distance. Under controlled conditions, the accuracy of this method can be very high. The literature review highlighted a recent study that validated OpenPose against a gold-standard motion capture system for Taekwondo athletes performing scripted movements [4]. In that a perfect controlled setting (multiple high-speed cameras, no occlusions or protective gear), the marker less pose system achieved root-mean-square errors of only 10 - 30 cm for 2D positions and showed excellent agreement (ICC > 0.90) with the mocap ground truth. This suggests that pose estimation is capable of near-manual (even near-marker) accuracy for tracking athlete's positions. Several instances were observed where the pose-estimated positions matched the manual annotations almost exactly. This shows the potential of pose-based tracking as a viable measurement tool, backing up early positive findings in controlled environments.

The pose model's performance declined in real competition footage, unlike a clean laboratory scene, the videos featured referees walking into view, audience members in the background, motion blur, and athletes wearing headguards and trunk protectors that obscured joints. Consequently, OpenPose occasionally produced extraneous or incorrect detections. Additionally, without an identity mechanism, the algorithm has no concept of which two people are the athletes of interest. During, heavy occlusion events (one athlete blocking the other, or limb entanglement during a clinch) caused pose results to degrade: keypoints from both athletes could become merged or assigned to the wrong individual.

This led to some frames where the pose tracker lost track of one athlete's leg or swapped the athletes' positions entirely, leading to spikes and dips in the measured distance.

Despite these challenges, when OpenPose had a clear view of both athletes, the output was very precise positional data, often capturing subtle shifts in stance or body lean that a bounding box cannot. This discovers a theme from literature: pose estimation provides richer information content than plain detection, enabling analyses of technique and biomechanics in addition to simple positions. In this context, pose data enabled an alternate calculation of distance (midpoint between feet, as per Maloney's definition) and could potentially allow automatic identification of actions (e.g., a sudden change in leg keypoint positions could indicate a kick attempt). In summary, pose estimation trials show both the promise and the pitfalls of marker less motion capture in a real competition setting. The results align with previous validations in that the accuracy can be high under ideal conditions.

9.5 Combining Bounding Boxes and Pose Estimation

Given the strengths of the bounding box for identification and pose estimation for detailed biomechanics, a hybrid system was developed that combines both. In this combined system, the bounding box detector first locates and labels each athlete and then the pose model focuses on those regions to extract the skeleton keypoints of the identified athletes. The rationale was that the detector's robust identity tracking could prevent the pose model from mislabelling extra people, while the pose model's precision could refine the positional measurement. The evaluation of this hybrid method showed it to be the best-performing solution overall: it reduced the incidence of identity swaps and false keypoint assignments and produced more stable distance measurements during occlusions than either method alone. Quantitatively, the hybrid approach achieved higher agreement with

ground-truth distances across all scenario categories (viewpoints, orientations, containment levels) than the standalone approaches.

The success of the approach is twofold. First, using bounding boxes for identity ensured that the pose algorithm only returns keypoints for the two fighters and ignores irrelevant people, a crucial improvement given OpenPose's tendency to pick up every person in the frame. Second, the pose data enhanced the spatial accuracy of the measurements; rather than relying on a rough torso-box centre, it is possible to calculate distance from more anatomically meaningful points (like hip or foot midpoints). The hybrid system thus produced distance metrics that coaches can trust more, because the identity is consistently correct and the measurement is tied to actual body landmarks. In essence, this approach brought together the reliability of object detection and the granularity of pose estimation.

It is worth noting that while the hybrid system markedly improved performance, it is not infallible. Extremely severe occlusions (e.g., one athlete completely behind the other from the camera's perspective) still reduced accuracy, as neither bounding boxes nor pose can create information that the camera does not see. Additionally, the processing pipeline becomes more complex, which in a real-time scenario could introduce latency (though this research study did not focus on real-time operation, this is a consideration for deployment of such systems). The research demonstrates a viable path forward by fusing specialised models, a strategy likely to be important in many applied settings. By validating this approach on authentic competition data, this research study provides a template that future systems (perhaps with even more advanced models) can build upon to achieve even higher reliability. The hybrid approach essentially allowed this study to reach closer to the manual annotation benchmark while maintaining practical automation, thus fulfilling a central aim of the research study, to see if CNNs can measure movement in elite Taekwondo. The answer, with this integrated method, is a cautious yes: it can be

done, with the caveat that careful system design (and not just off-the-shelf use of a single model) is necessary to handle the intricacies of real-world sport footage.

9.6 Measurement Challenges and Contextual Factors

One of the most valuable aspects of this research study was the scenario-based error analysis, which exposed where both operators and AI struggle in measuring Taekwondo movements. Chapter 4 introduced a framework breaking down scenarios by viewpoint, athlete orientation, and occlusion level (containment), later chapters utilised this to pinpoint performance drops. A primary challenge is occlusion, approximately 22% of frames in front or back-facing scenarios (where athletes aligned one behind the other relative to camera) resulted in “major containment” (>75% overlap of one athlete by the other). In these situations, both manual and automated methods saw notable accuracy degradation. For instance, in high containment frames automated distance errors were often double those of minimal containment frames, and operators showed larger disagreements as well. This confirms the intuition that occlusion is a fundamental obstacle, matching reports from other sports: multi-player tracking studies have long flagged occlusion as a leading source of error [104], [105]. In this research it was possible to quantify the level of error for Taekwondo, in one out of five exchanges from a front/back camera angle, you can expect severely reduced accuracy from any single-camera measurement system. A possible remedy is using multiple camera angles to cover blind spots, such as a recent study for basketball with multi-view tracking [90], but when evaluating using historical footage, this is not possible. Thus, acknowledging these danger zones of reliability is crucial for end-users; analysts must understand that data from those moments carry higher uncertainty.

Another key factor is camera perspective and calibration. The results demonstrated that the overhead viewpoint (when available) yielded the most consistent measurements in

both manual and automated. This is due to its orthogonal view of the mat, resulting in nearly true-to-scale distances. In contrast, the standard broadcast viewpoint (at an angle from one side) introduced perspective distortion: distances along the depth of the view were compressed, and without proper calibration this led to systematic underestimation of actual distances. This was addressed by applying a homography calibration to convert pixel coordinates to real-world coordinates, which improved accuracy, but even then, the variance in error was larger in the side-view videos. This finding echoes a previous analysis that showed that a shallow camera angle (30° from perpendicular) can increase positional measurement errors by 2.5 times when compared to a top-down 90° angle [92]. They also found that using an appropriate 2D calibration grid greatly improved accuracy versus a simplistic one-dimensional calibration. In this study, the benefits of a more detailed calibration by using multiple reference points on the mat to correct for perspective was used. Practically, this suggests that whenever possible, analysts should capture or use overhead footage for quantifying movements; if stuck with side footage, one must calibrate and accept a higher error margin. While CNN models handled the broadcast view reasonably well the viewpoint bias in accuracy was evident and must be factored into any deployment of such measurement systems.

This study also highlights sport-specific visual complexity as a challenge: the athletes wear similar (often identical white) uniforms and protective gear, which made both manual and automated identification harder. The operators in this study occasionally mis-tagged which athlete was which when they spun around, and the CNN had to be trained to recognise other cues such as headguard colour to distinguish them. The literature points out that false positives and identity confusion are common in such settings[106], for example, referees in white shirts were initially sometimes detected as athletes. Additionally, fast dynamic actions (spinning kicks, rapid footwork) introduced motion blur and temporary shape distortions that challenged the vision models. The error analysis

in Chapter 8 showed that frames during kicking actions had higher average error than frames during static footwork, correlating with the rapid distance closure and occlusion that kicks involve. This ties into the finding that distance tends to decrease sharply during kicks (by ~26% on average), exactly when measurement is hardest. In essence, the most critical moments (attacks) are the toughest to measure, a point that shows why achieving reliable automated tracking in combat sports is a non-trivial issue.

9.7 Implications for Applied Sports Science Analysis

Traditionally, coaches and sports scientists have relied on labour-intensive manual analysis or rudimentary metrics to gauge aspects like athlete distance, movement patterns, and work-rate. This research demonstrates that modern computer vision, specifically CNN-based tracking, can augment and in many cases revolutionise this process by providing objective, continuous data that was previously impractical to obtain in live or archival settings. One immediate implication is the ability to quantify and monitor the tactical use of distance with much greater precision. Coaches have long emphasised maintaining optimal distance for attack and defence, as evidenced by the common use of distance categories in training. With the validated system, a coach could, for example, review a match and see that their athlete on average kept 2.5 meters away from the opponent except during scoring exchanges where distance closed to 1.2 meters. More critically, these numbers would be backed by automated measurements rather than subjective estimation. This kind of insight can inform training: if an athlete consistently allows an opponent to enter a short range without initiating a counter, the coach now has hard data to reinforce that feedback.

Manually annotating thousands of frames is so time-consuming that studies in the past were limited to analysing a few selective moments during critical scoring events. With automation, now it is possible to process entire tournaments, generating datasets of

positioning and movement that allow for identifying patterns linked to success. For instance, do higher-ranked athletes maintain a longer average distance (playing a more evasive game) or a shorter one (engaging more)? Are there style differences between weight categories or between male and female competitors in terms of preferred engagement distance? Objective motion data like distance covered, and positioning are essential for understanding performance [1], and the system provides a means to gather that data in Taekwondo where previously it was not possible. The broader sports analytics community has seen similar transitions such as in football and basketball; optical tracking data has enabled detailed tactical analyses and new performance indicators. This aligns with previous future works suggestions of other studies were applying AI not just in training simulations but in match analysis to extract athletes' strengths and weaknesses [77].

Another implication lies in efficiency and objectivity of judging and coaching. There is ongoing discussion in combat sports about supplementing judges with objective data to reduce bias or error. While the system is not developed for real-time judging, it suggests that certain metrics (such as within scoring range) could be provided to commentators or analysts to enrich live broadcasts or post-fight breakdowns. Importantly, this data is now trustworthy: by benchmarking against manual and coach evaluations, and by ensuring the automated outputs are within a known error bound.

In summary, the successful application of CNN-based measurement in actual competition footage signals a shift for applied performance analysis in Taekwondo. The implications extend beyond Taekwondo: the approach can be a model for other combat sports (like karate, judo, boxing) where similar challenges exist. In a previous study using AI for boxing noted the importance of integrating kinematic data for classifying punches, which coincides with the focus on positional data for kicks and movement [107]. The convergence of sports and AI is happening, the work adds to the evidence that, when done

rigorously, these technologies can yield actionable competitive insights. Coaches armed with these tools can verify or debunk assumptions such as “Athlete A is keeping distance better than Athlete B”, now it is possible to check the numbers, leading to more tailored and effective coaching interventions. This ultimately contributes to the evolution of training methodology, making it more data driven. These systems are aids, not replacements, analyst expertise is still needed to interpret why an athlete maintained or failed to maintain distance, and how that fits into tactical context. The integration of objective measurement with expert interpretation is the future of performance analysis, and this study is a practical step in that direction for the sport of Taekwondo.

For GB Taekwondo to operationalise the methods validated in this research study, several practical steps are required. First, standardised video capture protocols should prioritise overhead or elevated camera positions (minimum 4-metre height, 1920×1080 resolution, 30 fps) at domestic competitions where placement can be controlled. Second, the competition area must be calibrated using the eight-vertex DLT method (Chapter 3) with validation of calibration accuracy against known mat dimensions. Third, footage should be processed using the hybrid bounding box and pose estimation pipeline (Chapter 8), outputting continuous distance measurements in metres alongside frame-by-frame athlete positions for the highest accuracy. To align these quantitative outputs with coaching practice, GB Taekwondo coaches should establish personalised distance category thresholds using the methodology in Chapter 6, enabling translation of metric distances into tactically meaningful classifications. During an initial pilot phase, analysts should cross-validate automated measurements against manual annotation to confirm performance within expected error bounds, with particular attention to the scenario-stratified accuracy expectations established in this research study. Finally, as annotated footage accumulates, the detection and pose models should be iteratively refined through transfer learning on GB Taekwondo-specific data, improving robustness to evolving

competition environments and venue characteristics. This implementation pathway enables the national programme to transition from research validation to operational deployment, extracting objective movement data that supports evidence-based coaching decisions across Olympic and World Championship cycles.

9.8 Future Research

While this research study has demonstrated the viability of CNN-based movement tracking in Taekwondo, it also showed several areas for future research and improvement. One clear need is enhancing the system's robustness to the identified challenges. For occlusions, exploring a multi-camera setup or depth sensors could dramatically improve tracking when one view is blocked. Recent developments in marker less motion capture suggest that fusing information from multiple vision sensors (RGB + depth, or multiple angles) can address out-of-plane movement issues and improve overall accuracy. For example, using a side camera in conjunction with a front camera could allow the system to always maintain visibility of both athletes. Future work could test a lightweight version of the tracker that operates on two synchronised video streams, essentially implementing a real-time 3D triangulation of athlete positions. This aligns with findings from the broader field that combining 2D pose estimates from multiple views yields more reliable 3D kinematics, which in turn could be projected back into each view for improved 2D tracking [108].

Testing the dataset collected on various other models would be a valuable extension, as such the dataset in this study has been prepared in a way that it can be used to train or evaluate new algorithms. Techniques such as having the system detect when it is in a major occlusion scenario then default to a conservative estimate or request operator confirmation. This approach may be practical for professional analysis settings, ensuring that critical errors are minimised.

There are also opportunities to generalise the approach beyond distance measurement. The framework and models could be adapted to measure other variables of interest in Taekwondo, such as kick execution speed or frequency of directional changes. By adding modules for action recognition (perhaps using the pose data to classify specific kick techniques, similar to boxing studies with punch classifications [91]), one could create a more comprehensive automated analysis tool. This would fully automate not only where the athletes are, but what they are doing. With the groundwork of the integrated system, adding an action classification layer on top would be a natural progression. Technical documentation from OpenPose and others suggests that recognising known poses or sequences (like a roundhouse kick motion) is feasible with additional training. Future researchers may leverage the dataset collected in this study (the largest of its kind in Taekwondo to date) to develop and validate such capabilities.

Finally, improving the user-friendliness and speed of the system will be important for uptake. Currently, high accuracy often comes at the cost of processing time, running pose estimation on high-resolution video can be slow. But as models become more efficient (e.g., Lightweight OpenPose, AlphaPose improvements, etc.), near real-time analysis is possible. As such a system where immediately after a match, coaches get a report generated by a tool based on the research, summarising distances, movement patterns, and possibly flagged moments. Achieving that kind of turnaround would transform how feedback is delivered in high-performance sport. Recent reviews and industry movements show a push towards bringing AI-driven analysis onto the field of play in various sports. This work serves as a proof of concept that it can be done for combat sports.

9.9 Conclusions

This research study set out to validate whether CNN-based computer vision techniques can measure elite Taekwondo athletes' movements. Through a staged approach, from manual benchmarking to automated detection, pose estimation, and hybrid integration, this study has shown that these tools are not only feasible but can achieve accuracy on par with domain experts under many conditions. The study critically examined where both manual and automated methods under-perform, those scenarios to known challenges documented in the literature, and addressed some of them by combining methods. The findings both reinforce prior research and extend it. For applied sports performance analysis, this work opens the door to data-driven insights that were previously out of reach in Taekwondo. Bringing objective spatial analytics to a martial art where timing and distance are paramount. The study has also highlighted how the results align with and contribute to the ongoing convergence of AI and sport, where objective measurements can enhance coaching, talent development, and even judging. This summarises that CNN-based measurement can be used for positional analysis in elite Taekwondo. This research study demonstrates that an approach preserving the contextual insight of expert operator while leveraging the scalability and consistency of automated methods has the potential to strengthen the way positional information is extracted and interpreted in Taekwondo performance analysis. By establishing a rigorous manual benchmark and evaluating CNN-based systems against real-world scenarios, this work provides a foundation for more objective and scalable measurement in the sport. The framework introduced here offers a structured basis for assessing automated methods that can support future advances in Taekwondo performance analysis.

References

[1] S. Roberts, G. Trewartha, and K. Stokes, “A comparison of time-motion analysis methods for field-based sports Submission type: Original Investigation.”

[2] M. A. Maloney, I. Renshaw, and D. Farrow, “The interpersonal dynamics of taekwondo fighting,” *Int. J. Perform. Anal. Sport*, vol. 21, no. 6, pp. 993–1003, 2021, doi: 10.1080/24748668.2021.1968660.

[3] P. Andrews, N. Borch, and M. Fjeld, “FootyVision: Multi-Object Tracking, Localisation, and Augmentation of Players and Ball in Football Video,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Apr. 2024, pp. 15–25. doi: 10.1145/3665026.3665029.

[4] L. dos Santos Banks, P. R. P. Santiago, R. da Silva Torres, D. C. X. de Oliveira, and F. A. Moura, “Accuracy of a markerless system to estimate the position of taekwondo athletes in an official combat area,” *Int. J. Perform. Anal. Sport*, vol. 24, no. 5, pp. 479–494, Sep. 2024, doi: 10.1080/24748668.2024.2321738.

[5] L. dos Santos Banks, P. R. P. Santiago, R. da Silva Torres, D. C. X. de Oliveira, and F. A. Moura, “Accuracy of a markerless system to estimate the position of taekwondo athletes in an official combat area,” *Int. J. Perform. Anal. Sport*, 2024, doi: 10.1080/24748668.2024.2321738.

[6] C. Park and T. Y. Kim, “Historical views on the origins of Korea’s Taekwondo,” *International Journal of the History of Sport*, vol. 33, no. 9, pp. 978–989, Jun. 2016, doi: 10.1080/09523367.2016.1233867.

[7] J. D. Ahn, S. ho Hong, and Y. K. Park, “The Historical and Cultural Identity of Taekwondo as a Traditional Korean Martial Art,” *Int. J. Hist. Sport*, vol. 26, no. 11, pp. 1716–1734, Sep. 2009, doi: 10.1080/09523360903132956.

[8] J. Y. Kim, *Taekwondo Textbook*. Seoul, Korea: O-Sung Publishing Company, 2006.

[9] J. , D. Eisenhart, “History of Taekwondo in the USA.” Accessed: Oct. 04, 2023. [Online]. Available: http://www.taekwondo-training.com/education/brief-history-of-tae-kwon-do/history-of-taekwondo-in-the-usa

[10] TheKoreaTimes, “Taekwondo Grandmaster Lectures at Yonsei Univ,” Jan. 10, 2008. Accessed: Oct. 04, 2023. [Online]. Available: http://www.koreatimes.co.kr/www/news/special/2009/11/178_17108.html

[11] The Seoul Times, “Grand Master Jhoon Rhee returns home to serve as Youngsan Univ.’s Chair Professor,” Sep. 2004. Accessed: Oct. 04, 2023. [Online]. Available: https://theseoultimes.com/ST/?url=/ST/db/read.php%3fidx=1012

[12] U. Moenig and M. Kim, “The origins of World Taekwondo (WT) forms or P’umsae,” *Ido Movement for Culture*, vol. 19, no. 3, pp. 1–10, 2019, doi: 10.14589/ido.19.3.1.

[13] World Taekwondo, “COMPETITION RULES & INTERPRETATION,” 2018.

[14] Y. Li, F. Van, Y. Zeng, and G. Wang, “BIOMECHANICAL ANALYSIS ON ROUNDHOUSE KICK IN TAEKWONDO,” 2005.

[15] J. W. Kim, M. S. Kwon, S. S. Yenuga, and Y. H. Kwon, “The effects of target distance on pivot hip, trunk, pelvis, and kicking leg kinematics in Taekwondo roundhouse kicks,” *Sports Biomech.*, vol. 9, no. 2, pp. 98–114, Jun. 2010, doi: 10.1080/14763141003799459.

[16] C. Falco *et al.*, “Influence of the distance in a roundhouse kick’s execution time and impact force in Taekwondo,” *J. Biomech.*, vol. 42, no. 3, pp. 242–248, Feb. 2009, doi: 10.1016/J.JBIOMECH.2008.10.041.

[17] C. Menescardi, C. Falco, A. Hernández-Mendo, and V. Morales-Sánchez, “Design, validation, and testing of an observational tool for technical and tactical analysis in the taekwondo competition at the 2016 Olympic games,” *Physiol. Behav.*, vol. 224, Oct. 2020, doi: 10.1016/j.physbeh.2020.112980.

[18] C. Menescardi, C. Falco, C. Ros, V. Morales-Sánchez, and A. Hernández-Mendo, “Technical-Tactical Actions Used to Score in Taekwondo: An Analysis of Two Medalists in Two Olympic Championships,” *Front. Psychol.*, vol. 10, Dec. 2019, doi: 10.3389/fpsyg.2019.02708.

[19] C. Menescardi, C. Falco, C. Ros, V. Morales-Sánchez, and A. Hernández-Mendo, “Development of a Taekwondo Combat Model Based on Markov Analysis,” *Front. Psychol.*, vol. 10, Oct. 2019, doi: 10.3389/fpsyg.2019.02188.

[20] C. Falco *et al.*, “Influence of the distance in a roundhouse kick’s execution time and impact force in Taekwondo,” *J. Biomech.*, vol. 42, no. 3, pp. 242–248, Feb. 2009, doi: 10.1016/j.jbiomech.2008.10.041.

[21] I. Estevan, O. A. ´ Lvarez, C. Falco, J. Molina-Garci´a, G. Garci´a, and I. Castillo, “IMPACT FORCE AND TIME ANALYSIS INFLUENCED BY EXECUTION DISTANCE IN A ROUNDHOUSE KICK TO THE HEAD IN TAEKWONDO.” [Online]. Available: www.nsca-jscr.org

[22] J. Headrick, K. Davids, I. Renshaw, D. Araújo, P. Passos, and O. Fernandes, “Proximity-to-goal as a constraint on patterns of behaviour in attacker–defender dyads in team games,” *J. Sports Sci.*, vol. 30, no. 3, pp. 247–253, 2012, doi: 10.1080/02640414.2011.640706.

[23] L. T. Bercades, “USING ECOLOGICAL DYNAMICS AND EXPERT KNOWLEDGE TO EXPLORE EXPERTISE-APPROPRIATE PRACTICE PEDAGOGIES FOR THE TAEKWONDO ROUNDHOUSE KICK,” 2022.

[24] Y.-C. Jiang, K.-T. Lai, C.-H. Hsieh, and M.-F. Lai, “Player Detection and Tracking in Broadcast Tennis Video,” in *Advances in Image and Video Technology*, T. Wada, F. Huang, and S. Lin, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 759–770.

[25] C. Mcinerney, “Determining spatio-temporal metrics that distinguish play outcomes in field hockey,” 2017.

[26] S. J. Edgecomb and K. I. Norton, “Comparison of global positioning and computer-based tracking systems for measuring player movement distance during Australian Football,” *J. Sci. Med. Sport*, vol. 9, no. 1–2, pp. 25–32, May 2006, doi: 10.1016/J.JSAMS.2006.01.003.

[27] M. A. Maloney and B. Exsci, “ENHANCING REPRESENTATIVE PRACTICE DESIGN THROUGH CONSDERATION OF AFFECTIVE AND SITUATIONAL CONSTRAINTS IN COMBAT SPORTS,” 2018.

[28] F. Tornello, L. Capranica, C. Minganti, S. Chiodo, G. Condello, and A. Tessitore, “Technical-tactical analysis of youth Olympic taekwondo combat,” *J. Strength Cond. Res.*, vol. 28, no. 4, pp. 1151–1157, 2014, doi: 10.1519/JSC.0000000000000255.

[29] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, “A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System,” *Sports Med. Open*, vol. 4, no. 1, pp. 1–15, Dec. 2018, doi: 10.1186/S40798-018-0139-Y/FIGURES/8.

[30] J. Windt, K. MacDonald, D. Taylor, B. D. Zumbo, B. C. Sporer, and D. T. Martin, "Current Concepts '"To Tech or Not to Tech?"' A Critical Decision-Making Framework for Implementing Technology in Sport," *J. Athl. Train.*, vol. 55, no. 9, pp. 902–910, 2020, doi: 10.4085/1062-6050-0540.19.

[31] K. A. Matsushigue, K. Hartmann, and E. Franchini, "Taekwondo: Physiological responses and match analysis," *J. Strength Cond. Res.*, vol. 23, no. 4, pp. 1112–1117, Jul. 2009, doi: 10.1519/JSC.0B013E3181A3C597.

[32] CVAT.ai Corporation, "Computer Vision Annotation Tool (CVAT)," Sep. 2022. [Online]. Available: https://github.com/opencv/cvat

[33] Kinovea, "Measuring Distances." Accessed: Jun. 16, 2025. [Online]. Available: https://www.kinovea.org/help/0.8.15/en/120.html#:~:text=Measuring%20distances%20,a%20line%20over%20the%20segment

[34] Dartfish, "Dartfish Advanced Video Analysis Software," 2004, *Switzerland*.

[35] E. Casolino *et al.*, "Technical and tactical analysis of youth taekwondo performance," *J. Strength Cond. Res.*, vol. 26, no. 6, pp. 1489–1495, Jun. 2012, doi: 10.1519/JSC.0B013E318231A66D.

[36] L. Rohner, C. R. Abbiss, W. Poon, and O. R. Barley, "Reliability of time-motion analysis in striking combat sports," *Sci. Sports*, vol. 39, no. 8, pp. 654–664, Jan. 2024, doi: 10.1016/j.scispo.2023.12.004.

[37] P. Krustrup and J. Bangsbo, "Physiological demands of top-class soccer refereeing in relation to physical capacity: effect of intense intermittent exercise training," *J. Sports Sci.*, vol. 19, no. 11, pp. 881–891, 2001, doi: 10.1080/026404101753113831.

[38] M. Weston, C. Castagna, F. M. Impellizzeri, E. Rampinini, and G. Abt, “Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates,” *J. Sci. Med. Sport*, vol. 10, no. 6, pp. 390–397, Dec. 2007, doi: 10.1016/j.jsams.2006.09.001.

[39] J. Benjamin, “Profiling the technical tactical components of performance in professional soccer,” 2006. [Online]. Available: http://cronfa.swan.ac.uk/Record/cronfa42307

[40] “Applications of GPS Technologies to Field Sports Robert J. Aughey,” 2011.

[41] J. W. Cui, Z. G. Li, H. Du, B. Y. Yan, and P. D. Lu, “Recognition of Upper Limb Action Intention Based on IMU,” *Sensors*, vol. 22, no. 5, Mar. 2022, doi: 10.3390/s22051954.

[42] Y. Zhao *et al.*, “Image expression of time series data of wearable IMU sensor and fusion classification of gymnastics action,” *Expert Syst. Appl.*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.121978.

[43] L. Pezenka and K. Wirth, “Reliability of a Low-Cost Inertial Measurement Unit (IMU) to Measure Punch and Kick Velocity,” *Sensors*, vol. 25, no. 2, Jan. 2025, doi: 10.3390/s25020307.

[44] C. Yu, T. Y. Huang, and H. P. Ma, “Motion Analysis of Football Kick Based on an IMU Sensor,” *Sensors*, vol. 22, no. 16, Aug. 2022, doi: 10.3390/s22166244.

[45] D. Mosler, M. Błażkiewicz, T. Góra, G. Bednarczuk, and J. Wąsik, “Using a long short-term memory model to predict force values of Taekwon-do turning based on spatio-temporal parameters,” *Acta Bioeng. Biomech.*, vol. 27, no. 1, 2025, doi: 10.37190/ABB-02565-2024-02.

[46] L. Needham, M. Evans, D. P. Cosker, and S. L. Colyer, “Can markerless pose estimation algorithms estimate 3d mass centre positions and velocities during linear sprinting activities?,” *Sensors*, vol. 21, no. 8, Apr. 2021, doi: 10.3390/s21082889.

[47] J. Lee, Y. Kim, and S. Nam, “Motion Capture based Taekwondo Motion Trajectory Visualization Media Art Technique,” *Journal of Digital Art Engineering and Multimedia*, vol. 6, no. 1, pp. 1–8, Jun. 2019, doi: 10.29056/idaem.2019.06.04.

[48] M. Radhakrishnan, R. Manikandan, R. Ramakrishnan, and R. Scholar, “Human Object Detection and Tracking using Background Subtraction for Sports Applications,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, 2013, [Online]. Available: https://www.researchgate.net/publication/276903638

[49] S. Gerke, S. Singh, A. Linnemann, and P. Ndjiki-Nya, “UNSUPERVISED COLOR CLASSIFIER TRAINING FOR SOCCER PLAYER DETECTION.” [Online]. Available: http://www.tosca-mp.eu

[50] A. Dearden, Y. Demiris, and O. Grau, “Tracking football player movement from a single moving camera using particle filters,” in *IET Conference Publications*, 2006, pp. 29–37. doi: 10.1049/cp:20061968.

[51] S. Baysal and P. Duygulu, “Sentioscope: A Soccer Player Tracking System Using Model Field Particles,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1350–1362, Jul. 2016, doi: 10.1109/TCSVT.2015.2455713.

[52] P. Garnier and T. Gregoir, “Evaluating Soccer Player: from Live Camera to Deep Reinforcement Learning,” Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.05388

[53] B. T. Naik and Md. F. Hashmi, “YOLOv3-SORT: detection and tracking player/ball in soccer sport,” *J. Electron. Imaging*, vol. 32, no. 1, p. 011003, Mar. 2022, doi: 10.1117/1.JEI.32.1.011003.

[54] J. Komorowski, G. Kurzejamski, and G. Sarwas, “FootAndBall: Integrated player and ball detector,” Dec. 2019, doi: 10.5220/0008916000470056.

[55] B. T. Naik and M. F. Hashmi, “Ball and Player Detection & Tracking in Soccer Videos Using Improved YOLOV3 Model,” Jun. 03, 2021. doi: 10.21203/rs.3.rs-438886/v1.

[56] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” May 2014, [Online]. Available: http://arxiv.org/abs/1405.0312

[57] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.02696

[58] G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics,” Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[59] W. Liu *et al.*, *SSD: Single Shot MultiBox Detector*, vol. 9905. in Lecture Notes in Computer Science, vol. 9905. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-46448-0.

[60] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” Aug. 2017, [Online]. Available: http://arxiv.org/abs/1708.02002

[61] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” [Online]. Available: http://image-net.org/challenges/LSVRC/2015/results

[62] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN.” [Online]. Available: https://github.com/

[63] Z. Cai and N. Vasconcelos, “Cascade R-CNN: Delving into High Quality Object Detection,” 2017. [Online]. Available: https://github.com/zhaoweicai/cascade-rcnn.

[64] E. Quinn and N. Corcoran, “The Automation of Computer Vision Applications for Real-Time Combat Sports Video Analysis.”

[65] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple Online and Realtime Tracking,” Feb. 2016, doi: 10.1109/ICIP.2016.7533003.

[66] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649. doi: 10.1109/ICIP.2017.8296962.

[67] Y. Zhang *et al.*, “ByteTrack: Multi-Object Tracking by Associating Every Detection Box,” Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.06864

[68] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, “BoT-SORT: Robust Associations Multi-Pedestrian Tracking,” Jun. 2022, Accessed: Sep. 06, 2023. [Online]. Available: https://arxiv.org/abs/2206.14651v2

[69] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.08008

[70] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional pose machines,” in *CVPR*, 2016.

[71] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand Keypoint Detection in Single Images using Multiview Bootstrapping,” in *CVPR*, 2017.

[72] 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 Trans. Pattern Anal. Mach. Intell.*, 2019.

[73] H.-S. Fang *et al.*, “AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time,” Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.03375

[74] A. Newell, K. Yang, and J. Deng, “Stacked Hourglass Networks for Human Pose Estimation,” Mar. 2016, [Online]. Available: http://arxiv.org/abs/1603.06937

[75] J. Wang *et al.*, “Deep High-Resolution Representation Learning for Visual Recognition,” Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.07919

[76] N. Pasaribu, E. Merry, K. Icasia, J. Eliezer, C.-W. Lin, and F. Setiawan, “Taekwondo Poomsae-3 Movement Identification by using CNN,” Scitepress, Dec. 2022, pp. 32–38. doi: 10.5220/0010744000003113.

[77] M. C. Shin, D. H. Lee, A. Chung, and Y. W. Kang, “When Taekwondo Meets Artificial Intelligence: The Development of Taekwondo,” Apr. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app14073093.

[78] M. A. Maloney, I. Renshaw, and D. Farrow, “The interpersonal dynamics of taekwondo fighting,” *Int. J. Perform. Anal. Sport*, vol. 21, no. 6, pp. 993–1003, 2021, doi: 10.1080/24748668.2021.1968660.

[79] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, “ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation.” [Online]. Available: https://github.com/ViTAE-Transformer/ViTPose.

[80] J. Mara, S. Morgan, K. Pumpa, and K. Thompson, “The accuracy and reliability of a new optical player tracking system for measuring displacement of soccer players,” *Int. J. Comput. Sci. Sport*, vol. 16, no. 3, pp. 175–184, Dec. 2017, doi: 10.1515/ijcss-2017-0013.

[81] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511811685.

[82] Y. I. Abdel-Aziz and H. M. Karara, “Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry,” *Photogramm. Eng. Remote Sensing*, vol. 81, no. 2, pp. 103–107, 2015, doi: 10.14358/PERS.81.2.103.

[83] J. Theiner and R. Ewerth, “TVCalib: Camera Calibration for Sports Field Registration in Soccer,” Oct. 2022, [Online]. Available: http://arxiv.org/abs/2207.11709

[84] F. Magera, T. Hoyoux, O. Barnich, and M. Van Droogenbroeck, “A Universal Protocol to Benchmark Camera Calibration for Sports,” Apr. 2024, doi: 10.1109/cvprw63382.2024.00338.

[85] M. Javadiha, C. Andujar, E. Lacasa, A. Ric, and A. Susin, “Estimating player positions from padel high-angle videos: Accuracy comparison of recent computer vision methods,” *Sensors*, vol. 21, no. 10, p. 3368, May 2021, doi: 10.3390/S21103368/S1.

[86] K. Jae-Ok and D. Voaklander, “Effects of Competition Rule Changes on the Incidence of Head Kicks and Possible Concussions in Taekwondo,” 2015. [Online]. Available: www.cjsportmed.com

[87] T. Jung and H. Park, “THE EFFECTS OF BACK-STEP FOOTWORK ON TAEKWONDO ROUNDHOUSE KICK FOR THE COUNTERATTACK,” 2020.

[88] M. Buric, M. Ivasic-Kos, and M. Pobar, “Player Tracking in Sports Videos,” in *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 2019, pp. 334–340. doi: 10.1109/CloudCom.2019.00058.

[89] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, “Learning to Track and Identify Players from Broadcast Sports Videos,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1704–1716, 2013, doi: 10.1109/TPAMI.2012.242.

[90] S. Tanikawa and N. Tagawa, “Player tracking using multi-viewpoint images in basketball analysis,” *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 5, pp. 813–820, 2020, doi: 10.5220/0009097408130820.

[91] S. Manoharan, J. Warburton, R. S. Hegde, R. Srinivasan, and B. Srinivasan, “An active machine learning framework for automatic boxing punch recognition and classification using upper limb kinematics,” *PLoS One*, vol. 20, no. 5, p. e0322490, May 2025, doi: 10.1371/JOURNAL.PONE.0322490.

[92] N. Shishov, K. Elabd, V. Komisar, H. Chong, and S. N. Robinovitch, “Accuracy of Kinovea software in estimating body segment movements during falls captured on standard video: Effects of fall direction, camera perspective and video calibration technique,” *PLoS One*, vol. 16, no. 10, p. e0258923, Oct. 2021, doi: 10.1371/JOURNAL.PONE.0258923.

[93] D. Wood, M. Reid, B. Elliot, J. Alderson, and A. Mian, “The expert eye? An inter-rater comparison of elite tennis serve kinematics and performance,” *J. Sports Sci.*, vol. 41, no. 19, pp. 1779–1786, Oct. 2023, doi: 10.1080/02640414.2023.2298102;WEBSITE:WEBSITE:TFOPB;PAGEGROUP:STRING:PUBLICATION.

[94] “What is Video Analysis in Sports? The Ultimate Guide - Catapult.” Accessed: Oct. 24, 2025. [Online]. Available: https://www.catapult.com/blog/what-is-sports-video-analysis

[95] “Revolutionizing Boxing Training with Computer Vision Technology | 2024.” Accessed: Oct. 24, 2025. [Online]. Available: https://www.rapidinnovation.io/post/computer-vision-in-boxing

[96] “How to Track Strength & Conditioning Progress in Combat Sports — Jason Lau | Performance Purpose - Strength & Conditioning.” Accessed: Oct. 24, 2025. [Online]. Available: https://www.performancepurpose.ca/article/monitoring-and-testing-in-combat-sports

[97] D. Rösch, M. G. Ströbele, D. Leyhr, S. J. Ibáñez, and O. Höner, “Performance Differences in Male Youth Basketball Players According to Selection Status and Playing Position: An Evaluation of the Basketball Learning and Performance Assessment Instrument,” *Front. Psychol.*, vol. 13, p. 859897, May 2022, doi: 10.3389/FPSYG.2022.859897/BIBTEX.

[98] L. Bruce, D. Dwyer, and A. Fox, “Reliability of live and video-based coding in netball using the NetballStats application,” *PLoS One*, vol. 17, no. 6, p. e0269330, Jun. 2022, doi: 10.1371/JOURNAL.PONE.0269330.

[99] G. Jin, “Player target tracking and detection in football game video using edge computing and deep learning,” *J. Supercomput.*, vol. 78, no. 7, pp. 9475–9491, 2022, doi: 10.1007/s11227-021-04274-6.

[100] G. Apollaro, P. V. Sarmet Moreira, T. Herrera-Valenzuela, E. Franchini, and C. Falcó, “Time-motion analysis of taekwondo matches in the Tokyo 2020 Olympic Games,” *Journal of Sports Medicine and Physical Fitness*, vol. 63, no. 9, pp. 964–973, Sep. 2023, doi: 10.23736/S0022-4707.23.14995-4.

[101] Y. Jia *et al.*, “A narrative review of deep learning applications in sports performance analysis: current practices, challenges, and future directions,” *BMC Sports Sci. Med. Rehabil.*, vol. 17, no. 1, pp. 1–20, Dec. 2025, doi: 10.1186/S13102-025-01294-0/FIGURES/13.

[102] D. T. Pham, N. T. T. Thuy, and L. Q. Tran, “Sportsort: overcoming challenges of multi-object tracking in sports through domain-specific features and out of view re-association,” *Mach. Vis. Appl.*, vol. 36, no. 6, Nov. 2025, doi: 10.1007/S00138-025-01756-Y.

[103] K. Vats, P. Walters, M. Fani, D. A. Clausi, and J. Zelek, “Player Tracking and Identification in Ice Hockey,” Dec. 2021, Accessed: Oct. 18, 2025. [Online]. Available: http://arxiv.org/abs/2110.03090

[104] A. Cioppa *et al.*, “SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos”, Accessed: Oct. 18, 2025. [Online]. Available: www.soccer-net.org.

[105] H.-T. Chen, C.-L. Chou, T.-S. Fu, S.-Y. Lee, and B.-S. P. Lin, “Recognizing tactic patterns in broadcast basketball video using player trajectory,” 2012, doi: 10.1016/j.jvcir.2012.06.003.

[106] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, “IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 1 Learning to Track and Identify Players from Broadcast Sports Videos”.

[107] S. Manoharan, J. Warburton, R. S. Hegde, R. Srinivasan, and B. Srinivasan, “An active machine learning framework for automatic boxing punch recognition and classification using upper limb kinematics,” *PLoS One*, vol. 20, no. 5, p. e0322490, May 2025, doi: 10.1371/JOURNAL.PONE.0322490.

[108] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand Keypoint Detection in Single Images using Multiview Bootstrapping”.