Submitted in fulfillment of the requirements for the degree of
Master of Science in Computer Science:
Data Science and Artificial Intelligence

Pose Estimation Analyses for Volleyball Players

Jakub Niedziela

Adviser: Dietrich Heiser
Prof. Tim Verdonck

Supervisor: Leonid Kholkine

IDLab
INTERNET & DATA LAB
University of Antwerp

Faculty of Science
2024

Contents

Acknowledgements

I am immensely grateful to my promotor, Leonid Kholkine, for his insightful guidance and unwavering support throughout the past year. His dedication and persistence in mentoring me have been crucial to my development and success in completing this thesis.

I would also like to extend my sincere thanks to Dietrich Heiser from i-LUDUS. His foundational work laid the cornerstone upon which I built my thesis. His expertise in the field and his passionate commitment to developing volleyball analysis tools were incredibly inspiring. Additionally, I am grateful for the enriching introductory courses he offered last summer, which were super useful in shaping my research.

A special thanks goes to my family, whose amazing support made it possible for me to study abroad in Antwerp. Their love and guidance have been my constant source of strength throughout my life.

Lastly, I want to thank all the wonderful people I have met in Antwerp. The friendships and connections I have made here have not only helped me navigate through the challenges of the master's program but have also made my time here very enjoyable. Discovering myself and enjoying life in this city would not have been the same without your presence.

To all, I am forever appreciative.

Summary

This thesis investigates the application of advanced AI techniques in the sport of volleyball, focusing on pose estimation and automated movement analysis to evaluate player attacks. By comparing the movements of training players against a reference model, the system aims to provide actionable feedback to coaches and athletes, helping them improve performance and possibly reduce injury risks. The literature revealed an absence of pose estimation-based systems designed for volleyball, which limits the ability to objectively analyse and improve player movements. That is why, the research developed an AI-driven system that used Body Part Embeddings deep learning model and phase detection to enable a granular analysis of specific body parts and different motion phases. Furthermore, it uses Dynamic Time Warping and cosine similarity scores for scoring final comparison of player and reference movements.

The findings demonstrate the system's potential effectiveness in providing detailed and objective performance feedback, highlighting its potential to improve volleyball training methodologies. Overall, this research contributes to the field of AI in sports, offering a tool for performance evaluation and laying the groundwork for future advancements in sports analytics. The study concludes with recommendations for further evaluation of the system, as well as proposes ideas for its expansion to other sports.

List of Abbreviations

AI Artificial Intelligence

BPE Body Part Embedding

CNN Convolutional Neural Network

DTW Dynamic Time Warping

FPS frames per second

IMU Inertial Measurement Units

PCA Principal Component Analysis

List of Figures

Chapter 1

Introduction

In recent years, the intersection of Artificial Intelligence (AI) and sports science has introduced new ways of improving athletic performance and coaching methodologies [4]. Among the many applications of AI in sports, pose estimation is particularly promising due to its ability to analyse and optimise athletes' movements in real time [5]. This thesis focuses on the sport of volleyball, where precision and technique play a very important role. It explores how advanced computer science algorithms can be used to improve volleyball training outcomes and its efficiency.

The main objective of this research is to develop a system, which by using a pose estimation and analysis methods on video data, analyses movements of volleyball players. By comparing the movements of a training player against a reference, the system provides feedback that can help coaches and athletes improve performance, while potentially minimising the risk of injury. This approach not only pushes the boundaries of traditional coaching techniques but also contributes to the growing field of AI in sports analytics.

This thesis explores the theoretical foundations of pose estimation and pose analysis; the development of a new algorithmic approach to movement analysis; and the practical use cases of this technology in sports and rehabilitation

training. Hence, the aim of this work is to bridge the gap between theoretical AI models and their real-world applications in sports. This will provide valuable tool for athletes and coaches to achieve optimal performance.

1.1 Background and Relevance

Volleyball is a complex sport that demands high levels of coordination, agility, and technique. It is therefore of no surprise that it could benefit from sophisticated training tools that can provide detailed and actionable feedback to players and trainers. Traditional methods of player training and performance analysis heavily rely on the subjective observation of coaches. This, while being very helpful, can benefit significantly from technological improvement.

Athlete training and performance evaluation methods have been deeply affected by the introduction of AI in sports [6]. One of the emerging technologies; pose estimation - a method that identifies the position and orientation of a person's body joints through images or video - proves to be especially important [5]. This technology has been affected by many advancements in recent years, mainly due to improvements in computer vision algorithms and the increasing availability of computational resources [7].

Nowadays, pose estimation is being widely used in various fields such as animation, gaming, and healthcare [8]. In sports, pose estimation analyses can provide quantitative and objective data, which proves helpful in analysing biomechanics of players' movement without need for specialised hardware. This is beneficial not only for performance improvement but also for injury prevention or rehabilitation, as it allows for precise correction of athletes' techniques [9, 10, 11, 12, 13].

In volleyball, the ability to analyse and replicate ideal movements can significantly impact a players' skill development [14]. The relevance of this research therefore lies in its use pose estimation and pose analysis techniques to compare a player's movements with a predefined reference movement. Such a comparison allows the user to identify and rectify potential mistakes, thus offering a usable tool which improves traditional coaching methods as well

as positively impacts player's development. Moreover, the integration of AI methods deals with an important problem in sports science - the lack of real-time and automated feedback mechanisms. By introducing a semi-automated system, this thesis aims to help trainers focus on other important aspects of training process like strategy or personal interactions.

Finally, the implications of this research can possibly go beyond volleyball training. It contributes to the broad field of AI by implementing a new semi-automated technique for complex movements comparison. This new method allows for more granular analysis and comparison of similar movements, improving already existing methods.

1.2 Problem Statement

Despite the rapid advancement in sport analytics technologies, there still exists a significant challenge in applying them effectively in volleyball training. Current training methods predominantly rely on subjective assessments from coaches, which lacks consistency and is limited by human observation capacity. Moreover, the adoption of video analysis tools - even though widely adopted in trainings - often lacks granularity necessary for a relevant biomechanical feedback. Such issue may negatively affect the training goal which is a detailed performance improvement.

The specific problem addressed by this work is the development of a reliable, objective, and automated system capable of analysing volleyball players' movements by comparing them to a reference movement. The comparison aims to detect deviations between two movements, providing a quantitative method of performance evaluation and improvement. The use of pose estimation and pose analysis technologies in volleyball faces several technical challenges, including:

1. **Accuracy and Precision:** Ensuring the pose estimation system accurately captures the complex movements, which can be fast-paced and involve multiple players simultaneously;

2. **Phase Detection:** Implementing an effective algorithm for detecting and segmenting different part of movements to allow for more detailed analysis and comparison of specific movement aspects;

3. **Real-Time Feedback:** Developing a system that not only analyses, but also provides real-time feedback to players and coaches during training sessions;

4. **Subjectivity in Evaluation:** Overcoming the subjective nature of movement quality assessment in volleyball. This issue might arise due to different opinions from coach to coach and dynamics of the movement;

5. **Data Integration and Processing:** Handling large volumes of data from various sources and processing this data efficiently while maintaining high levels of reliability.

The goal of this thesis is to address some of these challenges by developing an innovative AI-based system that improves the objectivity and effectiveness of volleyball training, utilising pose analysis techniques. This system is expected to improve the way volleyball training, player development and evaluation are performed by integrating state-of-the-art technological solutions into traditional methodologies. Therefore, this study aims to design, develop, and evaluate a novel AI-based system for pose evaluation that helps in the training and performance analysis of volleyball players. The specific objectives of this research are:

1. **Integrate Phase Detection in Movement Analysis:** Improve the system's capability by including an algorithm that detects and segments different movement phases within a volleyball attack. This segmentation will allow for a more detailed comparison and analysis of specific aspects of player movements against reference models;

2. **Provide Quantitative Feedback for Performance Improvement:** Generate objective, measurable feedback from the pose analysis that

can be used by coaches and players to identify areas in need of improvement. This includes comparing player movement with reference movement to identify differences between movements;

3. **Validate the System with Real-World Data:** Test the developed system using a comprehensive dataset of volleyball movements, ensuring that the system is robust, accurate, and applicable in real training environments.

Through these objectives, this research aims to significantly contribute to the use of AI in sports training. By providing a detailed, objective, and real-time analysis of athletic performance, the aim is to significantly improve traditional training methodologies. Furthermore, by open sourcing our solution we want to build the awareness of AI possibilities in sports as well as improve their availability to wider public.

1.3 Thesis Structure

The thesis consists of over several chapters, each dedicated to exploring different aspects of AI-based pose estimation and pose analysis for volleyball training. Chapter 2 begins with a comprehensive literature review that explores existing research on pose estimation in sports; applicable AI methodologies for motion analysis; and the current state of volleyball training technologies. This review establishes the key theoretical foundations for the following work and identifies gaps literature. In Chapter 3, the methodology of pose analysis system is described, including detailed overview of the algorithms and the final system architecture. It includes details regarding the integration of AI methods in the process and how issues like different frames per second (FPS) rates or variations in speed are being dealt with. Chapter 4 presents and interprets results of this research. Firstly, it describes the data collection process, which is then used for implementation and testing of proposed architecture. It also involves results gathered from real-world application scenarios providing deeper analysis of the outcomes. Chapter 5 concludes the thesis by summarising the contributions and significance of the findings, offering recommendations for future research and potential system improvements.

Chapter 2

Background and related work

This chapter explores the literature surrounding the application of pose estimation and pose analyses. It aims to show their implementation across various fields such as sports, physiotherapy, and other that require movement analysis. By discovering the evolution and application of these technologies, it aims to set the stage for introducing the specific methodologies used in this thesis. Technological advancements made in recent years are identified. Moreover, existing gaps in the literature that this thesis aims to fill are outlined.

The review begins with an overview of pose estimation techniques, presenting the reader with their development to their current state. This includes a discussion on how these methods evolved in last decades mainly due to rapid advancements in deep learning. The current state-of-the-art frameworks in modern 2D pose estimation and work on 3D pose estimation are mentioned.

In next part, the review presents use of machine learning technologies in sports. Firstly, the attention is given to use of wearable sensor devices. Although not particularly relevant to the video-based pose analysis, this part presents some interesting ideas how machine learning and AI can be integrated into movement analysis. Following wearable sensors the video-based methods are explored. These include application of machine learning for classification to detect different type of movements; to detect wrong movements; or even to suggest correct movements. Some of the works also focus

on rule-based assessment, which can be viewed as a combination of modern technologies and traditional approach. Moreover, the idea of pose evaluation to find small issues in movement is presented. Finally, studies focusing on direct comparison of two movements are explored.

In the key methodologies section, the review explores ideas most relevant to this research. These includes normalisation techniques, Dynamic Time Warping (DTW) and deep learning method. Works implementing these approaches are important to this thesis, as they present ideas to overcome challenges concerning accuracy and robustness of movement analysis in dynamic environments.

Finally this chapter discusses the gaps in the literature. This is the lack of specific applications in volleyball training as well as the need for methodologies regarding movement segmentation and comparison of complex movements. By highlighting these gaps, this review not only justifies the proposed research but also sets a clear path for subsequent chapters, to build on the existing knowledge and advance pose analysis in volleyball.

2.1 Pose Estimation

Rapid advancements in deep learning throughout the 2010s made way for significant progress in the field of pose estimation. This led to the development of new methodologies that have expanded capabilities and applications of pose estimation in sports and human movement analysis.

DeepPose [15] was among the first works to use deep neural networks to the task of human pose estimation, which marked a significant step from traditional model-based approaches. This method utilised convolutional neural networks (CNNs) to directly regress the coordinates of body joints from the full image input. By treating pose estimation as a regression problem, DeepPose improved the precision of pose predictions compared to earlier techniques. The use of CNNs and large datasets enabled the system to learn better feature representations, which is great for handling the variability in human poses across different scenarios and environments.

Following the advancements introduced by DeepPose, OpenPose [16] extended the capabilities of pose estimation to multi-person scenarios. This was an improvement from earlier systems, which were only able to handle images of individual objects. OpenPose introduced the concept of part affinity fields to map the connections between body parts, enabling effective detection of poses of multiple objects in the same image.

There are currently several state-of-the-art models in 2D pose estimation, each offering unique strengths. OpenPose [16] is known for its real-time, multi-person detection capabilities. AlphaPose [17] is optimised for crowded scenes, effectively distinguishing overlapping figures. DeepCut [18] provides comprehensive pose estimation by labeling body parts in dense environments, however it is computationally intensive. HRNet (High-Resolution Network) [19] maintains high-resolution, which in result improves the accuracy of keypoint detection. YOLO [20] models adapted for pose estimation are known for their speed and efficiency, ideal for real-time applications. Meta's Detectron2 [21] framework employs advanced neural network architectures to improve accuracy and speed in detecting human pose. Together, these models improve the capabilities and applications of 2D pose estimation technology, making them powerful tools for developing robust and scalable pose estimation solutions. Besides advancements in 2D pose estimation, some research also focuses on exploring 3D pose estimation from two-dimensional video inputs [22]. Such approaches utilise deep learning models to extrapolate three-dimensional information. This can offer a better understanding of human motion without need for specialised hardware like depth sensors or multi-camera setups.

2.2 Pose Analysis in Sports

This section explores current research, focusing on applications of pose estimation in sports analysis. It illustrates how pose analysis progress has impacted improvement of athletic performance, injury prevention and rehabilitation. The review covers range of methodologies, from mechanical models to the latest deep learning based approaches. By examining these

developments, following section establishes a solid foundation for understanding the significance of pose analysis in sports and its potential to transform traditional training and performance evaluation methods.

Some recent studies in sports science have explored use of wearable sensors for analysis of sports activities, especially in volleyball. Sensors – typically Inertial Measurement Units (IMU) – are equipped with accelerometers and gyroscopes that capture precise data about movements during gameplay. For instance, in beach volleyball, researchers have demonstrated use of a wrist-worn gyroscope to recognise and classify different serve types [23]. Another study explored use of sensors for evaluation of dominant and non-dominant hand movements during volleyball actions. The findings suggest that besides recognising specific volleyball actions; sensors also help in understanding biomechanical differences between players [24]. Moreover, use of deep learning techniques with sensor data, such as CNNs, has shown to enhance the classification accuracy of beach volleyball activities [25]. Volleyball skill assessment using a single wearable micro-IMU has been applied to detect different skill levels among players [26]. These works demonstrate how technological advancements can improve athletic performance. Although they do not use pose estimation methods, they show there is high interest in analysing volleyball motions.

Video-based pose estimation and classification represents another significant part of sports analysis. These approaches use machine learning models to analyse data extracted from videos using pose estimation algorithms. Because of this we can predict incorrect movements or even suggest how optimal execution should look like. These methods provide coaches and athletes with real-time feedback and sometimes actionable insights. For instance, "AI Coach" system integrates human pose estimation to identify "bad poses" in sport training videos. Moreover, it helps athletes improve their form by suggesting optimal pose sequences [10]. In combat sports, real-time video analysis system uses a computer vision to recognise different movements and generate statistics, helping in development of more effective training routines [27]. In basketball, a system developed for shooting prediction and pose correction utilises the OpenPose system for joint detection. It is then combined with K-Nearest Neighbor model and conditional Generative Adver-

sarial Networks for posture suggestion [28]. This approach not only predicts the outcomes of basketball shots but also suggests corrections to the athlete's shooting form. Some applications extend beyond sports into fitness and wellness, as demonstrated by a machine learning-based system for yoga training [11]. This system classifies yoga poses from video; detects incorrect postures; and provides real-time feedback to practitioners, helping them maintain correct posture during exercises.

Expanding on pose estimation for movement classification, some works focus on rule assessment in sports training and performance analysis. This involves recognising athletic movements and evaluating them with predefined; hard-coded standards to detect correct execution. For instance, in baseball swing movement was analysed to capture and assess hitters' poses in real-time [29]. The system measures limb angles and hip distances, followed by application of hard-coded rules that reflect expert coaching insights. Similarly, the analysis of simple exercise postures benefits from pose estimation by monitoring and correcting common movements such as squats and push-ups [12]. By comparing detected poses with ideal alignments of joints, which are defined by strict, hard-coded standards, the system provides immediate corrective feedback. These applications highlight the integration of rule-based assessments with video-based pose estimation, offering a tool for sports training improvement; ensuring users perform movements correctly and safely. The use of hard-coded rules enables consistent and reliable feedback, providing a standardised approach to training that is based on well-established sports science principles.

Several studies have improved the previous applications, by building solutions for detailed pose evaluation. These works use innovative algorithms to improve training process in many sports disciplines. One of them makes use of deep learning to assess golf swings against ideal models, scoring swing quality [30]. Besides delivering immediate feedback, this method also identifies specific areas in need of improvement. In another application, deep learning is used to provide visual feedback for golf training [31]. By analysing video sequences of golf swings, the system identifies deviations from optimal techniques and offers visual corrections. Additionally, the AI Golf tool proposes a new approach to analyse golf swings by suggesting intermedi-

ate motions for self-training [32]. It employs temporal CNNs to understand motion similarity, helping players to align their current executions with professional standards. Another example is a computer vision-driven system that improves training process by calculating differences in joint angles and comparing actions to standardised models [33]. All these methods present interesting approaches to the motion analysis in sports. They enable athletes to refine technique with quantifiable data; provide real-time feedback; and ensure correct movement execution.

In addition to methods that focus on individual motion analysis, several studies focus on direct comparison between two movements. This can considerably improve the ability to evaluate and correct human posture. One of such works integrates real-time feedback during workout sessions; using computer vision [34]. It is designed to analyse body posture by comparing it with reference model, identifying deviations in limb angles and providing corrective feedback. Another important work is creation of lightweight pose estimation models, adapted to provide real-time feedback on mobile devices [11]. This can improve the accessibility of motion analysis tools, making them available to a wider audience outside traditional gym environment. Moreover, applications such as Pose Trainer underscore the educational potential of pose estimation [13]. By utilising a pose estimation framework, they provide detailed and personalised recommendations for improvement of users' exercises. By using an interactive computer vision tool, the ExerciseCheck facilitates home-based physical therapy [35]. It supports a variety of exercises and provides real-time feedback by using RGB and RGB-D video inputs. Similarly, one study presents a tool to visualise user movements next to trainers which proves to be effective [36]. BalletNetTrainer introduces a new method for ballet training, using a combination of feature angle extraction and machine learning techniques [37]. By using the Dynamic Time Warping (DTW) to synchronise video frames, BalletNetTrainer provides actionable feedback to ballet dancers. The system highlights deviations from ideal poses and suggests corrections, improving training and preventing injuries because of wrong execution. Another work presents a system that adapts to various sources of human pose estimates for physical therapy applications [9]. Finally, development of a Body Part Embedding (BPE) model, improves the

understanding of 2D human motion similarity by decomposing body movements into distinct parts and assessing their similarities independently [1]. This approach uses a synthetic motion dataset along with real-world human annotations to train and validate its effectiveness. They also use a combination of new metrics to properly calculate movement embeddings. BPE significantly improves the correlation between computed motion similarities and human observations.

These articles provide an overview of the field of pose estimation and motion analysis, outlining recent innovations and advancements. They make solid foundations for further exploration, presenting a range of methodologies and applications from athletic performance analysis to therapeutic practices. The detailed examination of these technologies not only helps to understand current capabilities, but also creates ideas for future research. The literature review also reveals lack of pose estimation applications in volleyball. Furthermore, currently existing solutions in pose analysis field have not explored the idea of movement segmentation, which can be crucial for granular analysis in sports like volleyball. These gaps will be discussed in detail at the end of this chapter.

2.3 Key Technologies and Methods

In this section, key concepts from previously mentioned works are outlined, providing an overview of the methodologies used in modern motion analysis. This discussion provides the necessary technical background for this thesis. By connecting theoretical concepts to practical applications, this section introduces research that follows, ensuring a well-thought approach for the volleyball movement analysis.

Effective preprocessing and normalisation techniques play a crucial role in ensuring accuracy and robustness of models and analyses. Normalisation of pose data is an important preprocessing step, especially when considering variations in camera distance, angle, or the physical dimensions of users. Normalising the spatial coordinates of body parts ensures that pose analysis remains reliable irrespective of external factors such as the user's distance

from the camera or their body size [30, 13]. Other way to deal with these challenges is by using affine transformations which were utilised to adjust for variations in camera positioning and physical differences between users and reference models [34]. Furthermore, in real-time applications, particularly on mobile platforms, heatmap smoothing can play an important role. This technique involves averaging the heatmaps of consecutive frames to minimize sudden jumps in detected joint locations, providing a smoother and stable estimation of poses [38].

After the preprocessing steps, possible differences in movement start or speed need to be accounted for. DTW is a powerful algorithm widely used in the analysis of time series data, highly effective in situations where sequences vary in time or speed. DTW aligns sequences of data points by stretching or compressing them in time. Flexibility makes DTW an ideal tool for applications in motion analysis where the timing of actions can vary significantly between different attempts or individuals. DTW is used in plethora of studies on human motion analysis; being used to ensure correct movement frames are compared [9, 32, 34, 13, 35, 37, 38]. These works focus on employing DTW as a key component in systems designed to analyse and correct exercise posture, enhance physical therapy outcomes, or even in dance and sports training to ensure movements are performed correctly and effectively. They demonstrate how DTW's handling of temporal variations creates more advanced motion analysis and feedback mechanisms.

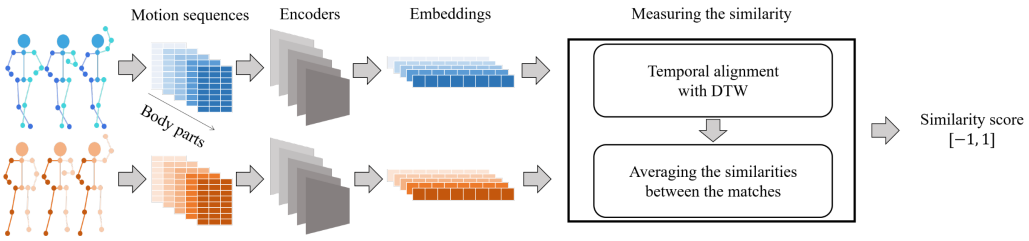


Figure 2.1: *Overview of Body Part Embedding method. The model produces embeddings from the input motion sequences, aligns them using Dynamic Time Warping and outputs a similarity score. Figure borrowed from [1].*

Introduction of BPE method offers a robust approach for movement comparison [1]. BPE essentially serves as a processing step, creating high-dimensional

embeddings of movements. These embeddings encode detailed body part dynamics. They can then be effectively used in following comparisons with use of DTW, aligning sequences of varying speeds or lengths. Such comparison focuses on the encoded vectors that reflect the fundamental structure of the movements. BPE approach uses strengths of neural networks to discover small differences in motion data, improving the effectiveness of traditional movement analysis methodologies. Overview of this method is presented on Fig 2.1. Notably, not only does BPE refine the input for better comparison but also improves the analytical capabilities of applications requiring detailed motion analysis, such as sports coaching or rehabilitation [1, 39].

For training purposes, BPE authors used SARA Dataset, which contains 4,428 3D motions which are then projected to 2D and used for model training. It contains many different 3D motions constructed using Mixamo software [40]. This dataset uses motion sequences from 18 different actors, which are then rendered into skeleton shape using Adobe Fuse software. Furthermore, it contains movements from different categories: Combat, Adventure, Sport, Dance; including many modifications of these movements. It is also known that BPE is a robust methodology for movement comparison as it was trained and evaluated by its authors, by calculating a correlation with similarity annotations on NTU RGB+D 120 dataset [41], achieving best results in comparison to other state-of-the-art methods.

Finally, all the mentioned methods in this section are useful to research conducted in this thesis. Starting with preprocessing and normalisation methods, followed by DTW and BPE; they will be described in detail in the Methodology chapter, laying foundations for the developed system.

2.4 Identifying Gaps

While reviewing the pose analysis literature, it is clear that not many studies focus on direct movement comparison. This gap is particularly visible if long, dynamic movements are considered. That is movements which involve a significant displacement of the person, such as volleyball attack. The literature here is underrepresented in comparison to more static motions like golf

swings, simple exercises, or rehabilitation movements. Moreover, most of existing research focuses on supervised methods, like classifying movements as correct/incorrect or providing corrective feedback based on predefined standards. Even though they are valuable, these approaches fundamentally differ from movement comparison, which is indeed an unsupervised task. Comparing dynamic movements of different people not only presents challenges in terms of methodological implementation but also in evaluation, as it requires a nuanced understanding of what constitutes similarity in dynamic human motion.

Another critical matter which is often overlooked in the literature is the impact of wrong keypoints detection. While many studies address this challenge to some extent, the problem remains related to the accuracy of pose estimation algorithms. Wrongly identified keypoints can lead to errors in movement analysis and comparison, affecting the reliability of such studies. Additionally, possible differences in video frame rates across samples were not addressed, even though they can critically affect the temporal alignment and comparison of movements. This issue is mainly present when analysing dynamic and longer motions, which may also naturally segment into different phases. This kind of segmentation adds more complexity to the analysis, as each phase might need to be handled differently.

Furthermore, the use of DTW in BPE [1] methodology has some shortfallings. The authors claim to use the DTW to align the embeddings and compute cosine similarity between sequences based on this alignment, however upon the inspection of the code the embeddings are treated more like a feature vector, not a time-series. Therefore, the DTW is not utilised to its full potential. This thesis will therefore try to address this gap.

These gaps show the need for deeper research regarding comparison of dynamic movements; improvement of pose estimation technologies; and development of methodologies able to handle complexities of human motion analysis. Addressing these challenges could lead to better tools for sports, rehabilitation, and other fields where insights into complex human movements are important. This research therefore aims to address and fill the gaps identified in the literature regarding the comparative analysis of longer and

dynamic movements. By focusing on more complex and displacement-heavy activities, this study aims to develop methodology that can effectively handle the challenges of comparing dynamic human movements. This includes facing the unsupervised nature of movement comparison and enriching the analysis process to accommodate for multiple phases in longer movements. Additionally, this research explores a solution to deal with potential inconsistencies caused by variations in video frame rates. Through these efforts, the study contributes to the field of motion analysis in volleyball, providing new tools and insights for evaluation of dynamic activities in a more detailed and accurate way. Even though this study focuses specifically on sport of volleyball, we believe there is a potential of applying developed algorithms to other sports in future.

Chapter 3

System Design

Following chapter introduces the reader to different methodologies used to analyse volleyball player movements in this thesis. The goal is to describe development of an algorithm that takes input of two videos and outputs a comparison score. First video is the reference video, while second video is the compared video which the user wants to evaluate. Subsequent text provides a deep dive into all the aspects of proposed design.

The system design for this study revolves around the development and implementation of a phased comparison approach for analysing volleyball player movements. The design uses advanced machine learning methodologies and incorporates a deep learning approach to align and compare the movements of a training player against a predefined reference movement. The overall flow of this process is depicted in the Fig. 3.1.

The initial phase of the design involves two video inputs: a reference video and a comparison video. The flow begins with the extraction of key joint positions (keypoints) from both videos. This extraction is performed using a pre-existing pose detection pipeline, developed by i-LUDUS, which uses the Detectron2 [21] as the pose estimation model. The algorithm identifies the keypoints for each joint in the body, and outputs a dataset that is used in further analysis. The use of an already established pose detection pipeline ensures consistency and reliability in the extraction process.

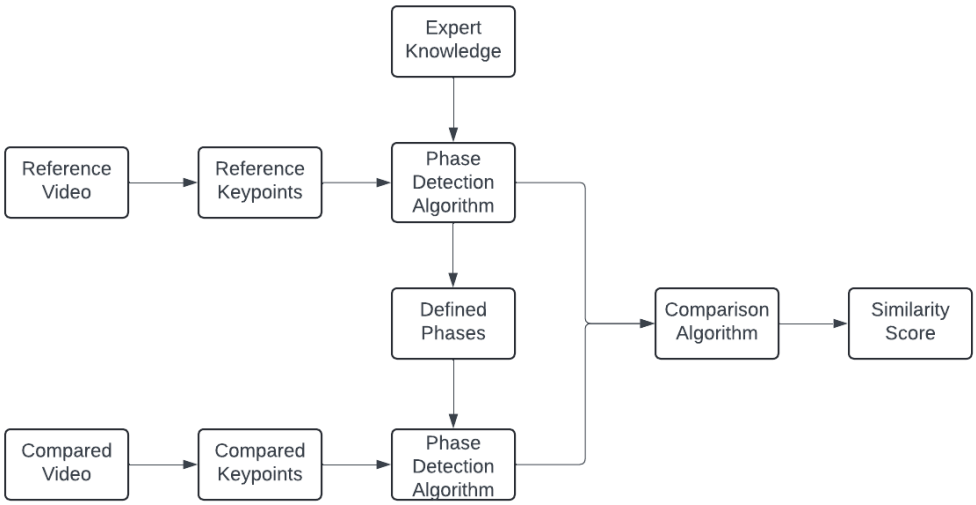


Figure 3.1: *Flow diagram representing system design. The keypoints extracted from videos are fed to phase detection algorithm, which also includes the expert knowledge. Resulting segments are fed into comparison algorithm which outputs similarity scores for each of detected phases.*

Following keypoints extraction, the next stage in the system design is phase detection. The phase detection algorithm is first applied to the keypoints data of the reference video. It explores different potential phase configurations, presenting options for the number of phases into which the movement can be divided. An expert, typically a coach or a user familiar with the sport, provides an important input by selecting the optimal number of phases based on their knowledge and experience. This expert input serves as a form of domain-specific knowledge to guide the segmentation algorithm, making sure the number of phases is meaningful.

Following the expert-driven segmentation of the reference video, the system proceeds to segment the compared video. The phase segmentation algorithm is applied to the normalised keypoints data of the compared video, using the number of phases determined by the expert in previous step. Because of that, consistency in phase segmentation approach is ensured; making it possible to analyse videos in phase-by-phase manner. Segmentation process therefore aligns two motions facilitating a granular comparative analysis.

Now that the videos are segmented into defined phases, next step is the com-

parative analysis stage. The comparison algorithm utilises a deep learning model based on BPE [1] to evaluate the movements within each phase. This model focuses on five body parts: right arm, left arm, right leg, left leg and the torso; to provide more detailed feedback. By encoding the movements of each body part and aligning them using DTW the algorithm assesses the similarity between corresponding phases in the reference and compared videos. The use of deep learning at this stage enables the system to capture complex patterns in the movements, providing robust comparison mechanism. Additionally, the model accounts for variations in body size and movement speed among different players, making the comparison better.

The final output of the system is a set of similarity scores for each body part in each phase. These scores provide a granular comparison between the reference and compared videos, highlighting areas of similarity and differences. The comparison approach with use of phases, guided by expert input and using advanced AI techniques, represents a robust methodology for evaluating volleyball player movements.

Following sections will closely examine the algorithmic details of the system. This includes an in-depth look at the preprocessing steps, phase detection, the implementation of BPE, and the use of DTW for movement alignment. Each section will focus on the technical aspects and methodologies used, providing an overview of how these components contribute to the overall system.

3.1 Preprocessing

Normalisation

In order to account for differences in video resolution and align keypoints within bounding box; a normalisation process is applied. This adjustment allows for consistent comparisons across videos of varying sizes and resolutions. The normalisation process is mathematically described as follows:

Given keypoints x and y, and the bounding box defined by the minimum

and maximum values of all the keypoints across entire motion:

$$\begin{aligned} x_{\min} &= \min(x), \quad x_{\max} = \max(x), \\ y_{\min} &= \min(y), \quad y_{\max} = \max(y), \end{aligned} \tag{3.1}$$

the normalised keypoints x_{norm} and y_{norm} are calculated using the formulas:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3.2}$$

$$y_{\text{norm}} = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \tag{3.3}$$

This process is applied for all the 15 joints to both x and y coordinates. This ensures that all keypoints fall within a range of [0, 1] for both the x and y coordinates, effectively normalising them relative to the full movement bounding box. Figure 3.2 presents application of this process. Top plot depicts two movements before normalisation, while bottom plot contains normalised skeleton motions.

Frame rate aggregation

Furthermore, to align the frame rates between the reference and compared videos, the system uses an aggregation process that adjusts the keypoints to match the target frame rate. The FPS aggregation process is mathematically described as follows:

Given a vector v representing some keypoints the aggregation factor k is defined as:

$$k = \frac{\text{original\_fps}}{\text{target\_fps}} \tag{3.4}$$

where $original\_fps$ and $target\_fps$ are the frame rates of the original and target videos, respectively. The function aggregates the vector v such that:

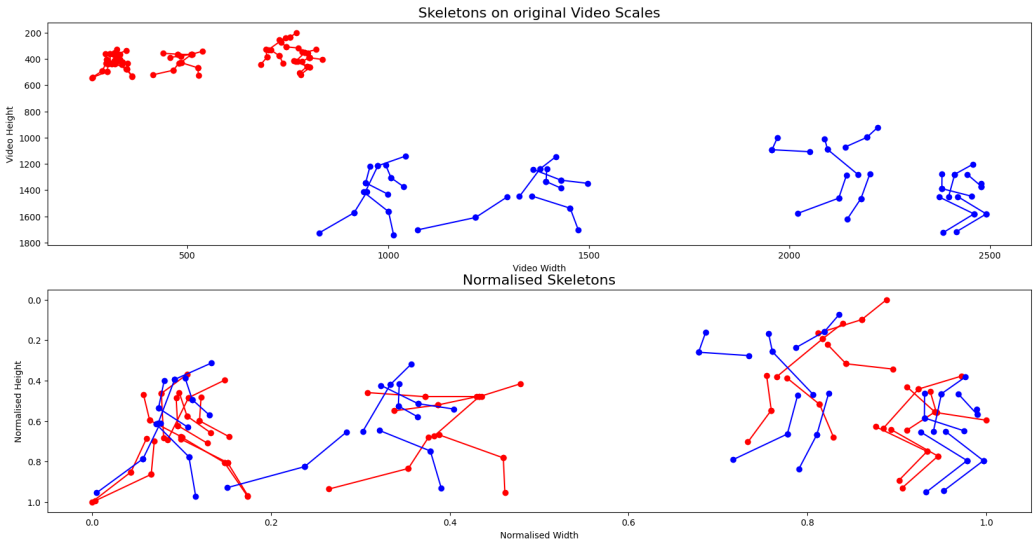


Figure 3.2: *Skeletons before and after normalisation. Red skeleton is from video with resolution of 1280 x 720 (720p). Blue skeleton is from video with resolution of 3840 x 2160 (2160 p).*

$$v_{\text{agg}}[i] = \frac{1}{k} \sum_{j=0}^{k-1} v[i \cdot k + j], \tag{3.5}$$

resulting in a vector v_{agg} where each value represents the average of k consecutive frames from the original vector v. This process ensures that the frame rate of the data aligns with the desired target frame rate. Fig. 3.3 presents an example of scaling y coordinates of right elbow from 60 to 30 fps.

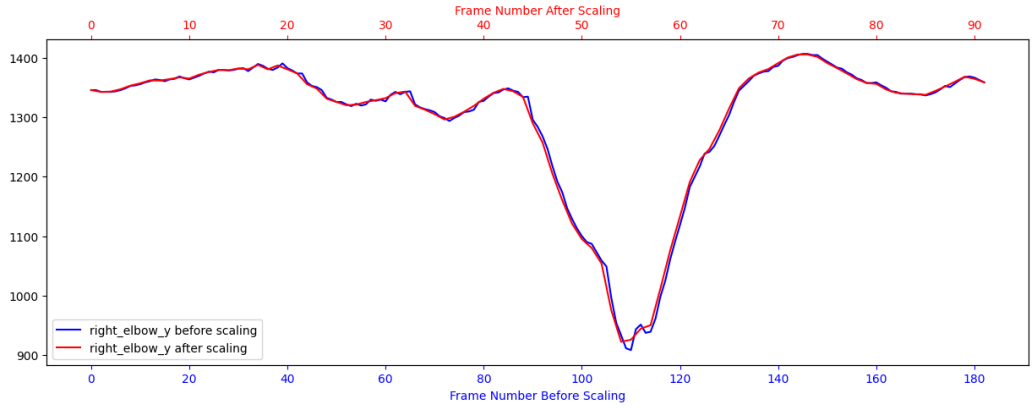


Figure 3.3: *Right elbow y coordinates before and after scaling from 60 to 30 fps.*

3.2 Phase Detection

The Kernel Change Detection algorithm, implemented in the ruptures Python package as 'KernelCPD' method [42], is useful for analysing data with many features. This makes it perfect tool for the phase segmentation of volleyball movements based on multiple keypoints. The algorithm works well with multivariate data because it considers each feature's contribution to changes in the overall statistical properties of the data. It identifies change points in a multivariate time series $Y = \{y_1, y_2, \ldots, y_T\}$. This section provides an overview of the kernel change point detection method, and how it is utilised on players keypoints data.

Given a $\mathbb{R}^d$-valued signal $y = \{y_0, y_1, \ldots, y_{T-1}\}$ with T samples. This signal is mapped onto a reproducing kernel Hilbert space $\mathcal{H}$, which is associated with a user-defined kernel function $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. The mapping function $\phi : \mathbb{R}^d \to \mathcal{H}$ is implicitly defined by $\phi(y_t) = k(y_t, \cdot) \in \mathcal{H}$. This leads to the following inner product and norm:

$$\begin{aligned} \langle \phi(y_s) \mid \phi(y_t) \rangle_{\mathcal{H}} &= k(y_s, y_t) \\ \|\phi(y_t)\|_{\mathcal{H}}^2 &= k(y_t, y_t) \end{aligned} \tag{3.6}$$

for any samples $y_s, y_t \in \mathbb{R}^d$.

The goal of kernel change point detection is to identify mean shifts in the mapped signal $\phi(y)$ by minimising the following objective function $V(\cdot)$:

$$V(t_1, \ldots, t_K) := \sum_{k=0}^{K} \sum_{t=t_k}^{t_{k+1}-1} \left\| \phi(y_t) - \mu_{t_k:t_{k+1}} \right\|_{\mathcal{H}}^2 \tag{3.7}$$

where $\mu_{t_k:t_{k+1}}$ is the empirical mean of the sub-signal $\phi(y_{t_k}), \phi(y_{t_k+1}), \ldots, \phi(y_{t_{k+1}-1})$, and $t_1, t_2, \ldots, t_K$ are the change point indices in increasing order (with the convention $t_0 = 0$ and $t_{K+1} = T$).

If the number of changes K is known beforehand, the optimal change points are determined by solving the following optimisation problem over all possible change positions $t_1 < t_2 < \cdots < t_K$ [43]:

$$\hat{t}_1, \ldots, \hat{t}_K := \arg \min_{t_1,\ldots,t_K} V(t_1, \ldots, t_K). \tag{3.8}$$

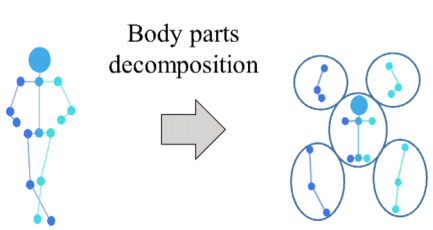
By minimising the above objective function, we can effectively detect change points in the signal, allowing for better analysis and understanding of the underlying data.

This method is applied to 15 normalised joint points to be consistent with other parts of system. Firstly, it is applied to the reference video and then the user is presented with a choice. The choice contains different possibilities of number of breakpoints in the movement (breakpoints segment motion into different phases). When the user chooses a value, this is remembered by the system and applied to the compared video. In the first step this is essentially the problem described by equation (3.8), with different values of K presented to the user. In the second step the problem formula is the same, however algorithm only uses the K value chosen by user. Finally, the keypoints in each of the detected phases are passed to BPE, which is described in the following section.

3.3 Body Part Embeddings

The BPE [1] model is the one of the main components of system used for analysing volleyball player movements. It operates by taking a sequence of joint coordinates and producing embeddings for each body part (Fig. 3.4). Embeddings are then used to assess the similarity between different motions. The BPE model architecture is designed to separately process the motion, skeleton, and camera view attributes of each body part through a series of specialised encoders and decoders, as depicted in Figure 3.5

The BPE model uses a modular architecture, where each of body parts: right arm, left arm, right leg, left leg, and torso has its own set of encoders and



Body parts	Composition joints
Right Arm (RA)	Right shoulder, right elbow, right wrist
Left Arm (LA)	Left shoulder, left elbow, left wrist
Right Leg (RL)	Right hip, right knee, right ankle
Left Leg (LL)	Left hip, left knee, left ankle
Torso	Nose, neck, right shoulder, left shoulder, right hip, left hip

Figure 3.4: *Joints decomposition used in Body Part Embedding. The detected keypoints are grouped into 5 body parts, using specific joints for each body part. Based on this aggregation, the final embeddings are created. Figure borrowed from [1].*

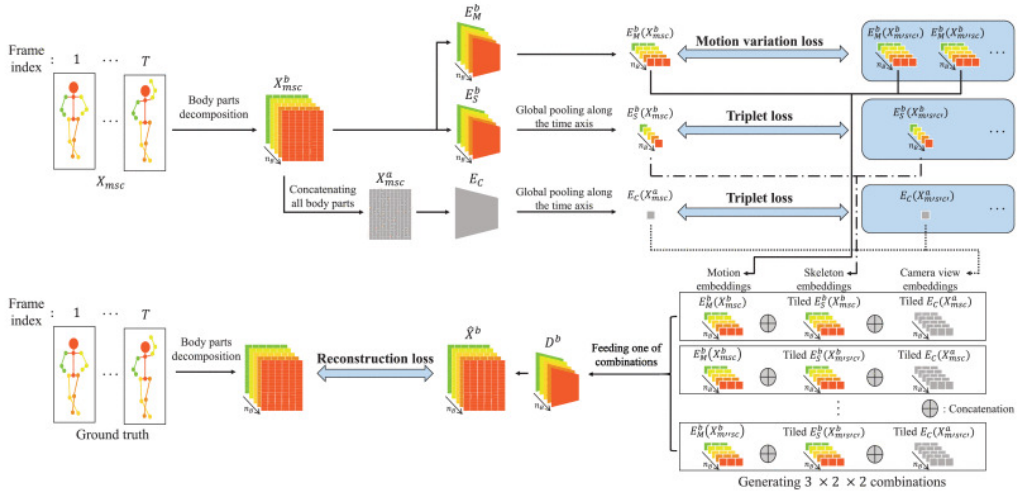


Figure 3.5: *Body Part Embedding model architecture. Each body part is drawn in a different color. The input is a sequence of frames representing different body parts, which are decomposed and processed through body part motion and skeleton encoders. Motion embeddings (E_M^b) and skeleton embeddings (E_S^b) are generated for each body part and pooled globally along the time axis. The motion variation loss and triplet loss are applied to the embeddings. A global pooling operation along the time axis is performed for camera view embeddings (E_C). The model reconstructs the body part sequences and computes reconstruction loss by comparing the generated sequences with the ground truth. Motion, skeleton, and camera view embeddings are concatenated in various combinations to generate embeddings. The bottom section of the figure shows the details of embedding generation and reconstruction processes. Figure borrowed from [1].*

decoders. This design allows the model to capture movement patterns of each mentioned body part independently. The structure is made of the following components:

1. **Motion Encoder (3.6a):** Each motion encoder processes a sequence of joint coordinates for the corresponding body part and produces a motion embedding. These embeddings capture the movement patterns of the body part over time. The encoder is built by stacking three layers of convolutional layers with batch normalisation and Leaky ReLU activation function.

2. **Skeleton Encoder (3.6b):** Each skeleton encoder also processes the joint coordinates but focuses on extracting the skeletal structure of the body part. The resulting skeleton embedding compresses the temporal information using global max pooling. This encoder helps to distinguish different body structures or poses for the same motion.

3. **Camera View Encoder (3.6c):** The camera view encoder takes the concatenated coordinates of all body parts and creates an embedding that accounts for the camera view during the recording. This ensures that variations in the viewing angle do not interfere with the motion similarity analysis.

4. **Decoders (3.6d):** The embeddings from the motion, skeleton, and camera view encoders are combined and fed into a body part decoder. Encoder then reconstructs the motion sequence for the specific body part. This process ensures that the combined features can effectively describe the body part's movement. The decoder is built by stacking two layers of upsampling, convolutional layer, dropout and Leaky ReLU activation function.

Furthermore, the BPE model employs several loss functions to train the encoder-decoder network, each focusing on different aspect of the motion data. Different loss functions are important aspect of model's ability to produce accurate and meaningful embeddings, which are then used for similarity comparison.

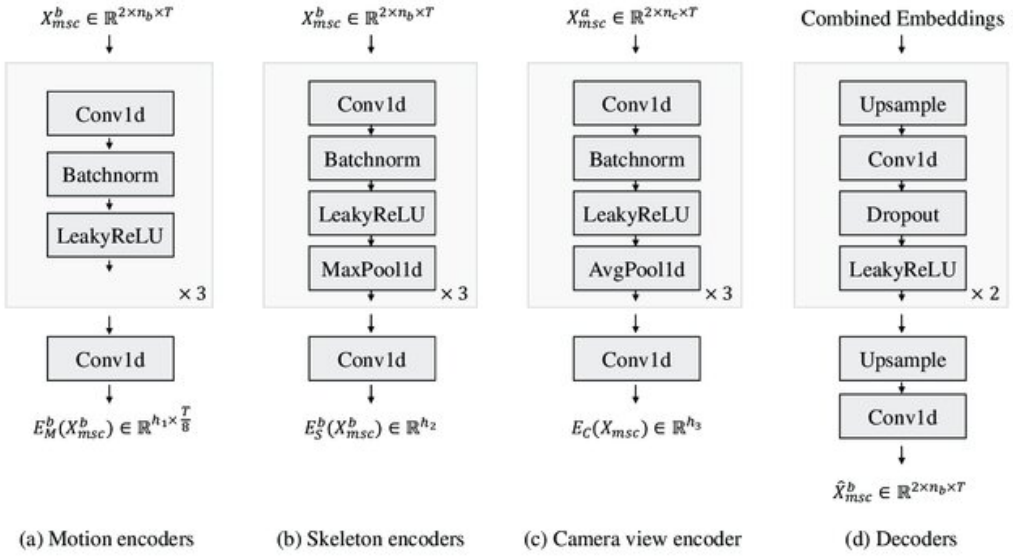


Figure 3.6: *Body Part Embedding Encoders and Decoders. (a) Motion Encoders: These encoders take a 2D sequence of a body part as input and pass it through layers of 1D convolutions, batch normalization, and Leaky ReLU activations, followed by another convolutional layer to produce motion embeddings (E^b_M). (b) Skeleton Encoders: Like motion encoders, these also process the 2D sequence of a body part but include max pooling layers to compress temporal information, resulting in skeleton embeddings (E^b_S). (c) Camera View Encoder: This encoder processes concatenated sequences of all body parts, using average pooling to generate camera view embeddings (E_C). (d) Decoders: The decoders reconstruct the input sequences from the combined embeddings. They include upsampling, convolutional layers, dropout, and Leaky ReLU activations to produce the final reconstructed output ($\hat{X}^b_{msc}$). Figure borrowed from [1].*

In order to defines loss function, the BPE authors first provide necessary mathematical definitions. Given M, S, C as sets of motion, skeleton and camera view attributes in training set, $M = \{m, m', m''\}$, $S = \{s, s'\}$, and $C = \{c, c'\}$. m and m' are from the same motion class with different characteristics, and m' is a motion from a different class (if m is from class low jump, then m'' is a high jump and m' is sitting). s and s' represent two skeletons with different body structures, and c) and c' are view angles. The motion sequence is then a combination of elements from M, S and C. They define set of motion sequences as $X = \{X_{ijk} \in \mathbb{R}^{2\times J\times T} \mid i \in M, j \in S, k \in C\}$. For example, X_{msc} and $X_{ms'c}$ are sequences representing the same motion m, with different skeletons and the same camera view. Furthermore, a set B is defined with $B = \{RightArm, LeftArm, RightLeg, LeftLeg, Torso\}$. It is composed of $n_B = 5$ body parts, to decompose a skeleton and construct body part embeddings. The motion sequence X_{msc} can be decomposed into specific body parts $X^b_{msc} \in \mathbb{R}^{2\times n_B\times T}$. Finally, $X^a_{msc} \in \mathbb{R}^{2\times n_c\times T}$ is given, where n_c is the sum of joints, given all body parts. Then, the loss functions are as follows:

- The motion variation loss is an upgraded version of the triplet loss, which includes information about similarity between samples. Triplet loss makes sure that similar samples are close in latent space, while different samples are far. It ensures that small variations in movement are reflected in the embeddings. The formula for the motion variation is:

$$\mathrm{var}(m, m'') = \frac{\|v_m - v_{m''}\|_1}{2 \times n_{v_m}}, \tag{3.9}$$

  where v_m and $v_{m''}$ are the characteristic vectors of the motions m and m'' (same motion class, but different characteristics), and n_{v_m} is the number of variables in the characteristic vector. The motion variation loss is then defined as:

$$\begin{aligned}\mathcal{L}^b_{\text{var}}\left(X^b_{msc}, X^b_{m's'c'}, X^b_{m''sc}\right) = \mathcal{L}^b_M\left(X^b_{msc}, X^b_{m's'c'}\right) \\ + \mathcal{L}^b_M\left(X^b_{m''sc}, X^b_{m's'c'}\right) \\ + \alpha\left\{d\left(z^b_{msc}, z^b_{m''sc}\right) - \beta \cdot \text{var}(m, m'')\right\}^2,\end{aligned} \tag{3.10}$$

where L^b_M is a triplet loss; d is a distance metric, and α and β are hyperparameters set to 1 and 0.1; and z^b_{msc} is an embedding of motion sequence X^b_{msc}.

- The skeleton embeddings triplet loss focuses on the skeletal structure of the body parts. It is defined as:

$$\mathcal{L}^b_S(X^b_{msc}, X^b_{m's'c'}) = \left[d(z^b_{ms'c'}, z^b_{msc}) - d(z^b_{ms'c'}, z^b_{m's'c'}) + \delta\right]_+ \tag{3.11}$$

where z^b_{msc} represents the skeleton embedding, obtained from X^b_{msc}.

- The camera view embeddings triplet loss focuses on the camera view aspect of the body parts, ensuring that variations in viewing angles do not interfere with the motion similarity analysis. It is defined as:

$$\mathcal{L}_C(X^a_{msc}, X^a_{m's'c'}) = \left[d(z^a_{ms'c'}, z^a_{msc}) - d(z^a_{ms'c'}, z^a_{m's'c'}) + \delta\right]_+ \tag{3.12}$$

where z^a_{msc} represents the camera view embedding, obtained from X^a_{msc} - a motion sequence of concatenated body parts.

- The reconstruction loss ensures that the reconstructed motion sequences accurately reflect the input data, maintaining the integrity of the motion and skeleton information. The formula for the reconstruction loss is:

$$\mathcal{L}^b_{rec} = \frac{1}{12}\sum_{i\in M}\sum_{j\in S}\sum_{k\in C}\left(X^b_{ijk} - \hat{X}^b_{ijk}\right)^2, \tag{3.13}$$

where X^b_{ijk} is the ground truth sequence and $\hat{X}^b_{ijk}$ is the reconstructed sequence for body part b. This loss helps the model disentangle motion, skeleton, and camera view information.

- The motion variation loss, skeleton embeddings triplet loss, and camera view embeddings triplet loss are combined to create the similarity loss term:

$$\mathcal{L}_{sim} = \sum_{b \in B} \mathcal{L}^b_{var} + \sum_{b \in B} \mathcal{L}^b_S + \mathcal{L}_C. \tag{3.14}$$

Finally, the total loss used to train the BPE model is a weighted sum of the individual losses:

$$\mathcal{L} = \lambda_1 \sum_{b \in B} \mathcal{L}^b_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{sim}} + \lambda_3 \mathcal{L}_{\text{foot}}, \tag{3.15}$$

where the parameters λ_1, λ_2, and λ_3 were set by authors to 1, 1, and 0.5, respectively. The $\mathcal{L}_{\text{foot}}$ is a foot velocity loss, presented in [44]. It prevents a phenomenon known as foot skating, which can cause significant errors in the motion of hands and feet. This combined loss ensures that the model learns to produce accurate motion embeddings while also maintaining realistic foot movements and capturing small motion variations.

The BPE model's architecture is designed to provide detailed comparisons of human movements by focusing on individual body parts. By separating the movement, skeletal structure, and camera view components, the model can capture the specific characteristics of each body part's motion in a better way. This modular approach ensures that small differences in movement are detected, while accounting for variations in body structure and viewing angles.

3.4 Dynamic Time Warping

After all the preprocessing, phase detection, and creation of embeddings DTW is used in the analysis, to align the embeddings from each of computed

phases.

DTW is an algorithm designed for aligning and comparing sequences that vary in time or speed. It works by constructing a cost matrix C where each element $c_{i,j}$ measures the distance between points x_i from one sequence and y_j from another. The goal of DTW is to find the path through this matrix that minimises the total distance. This aligns the sequences optimally despite differences in their timing or length. The process is visualised on Fig. 3.7.

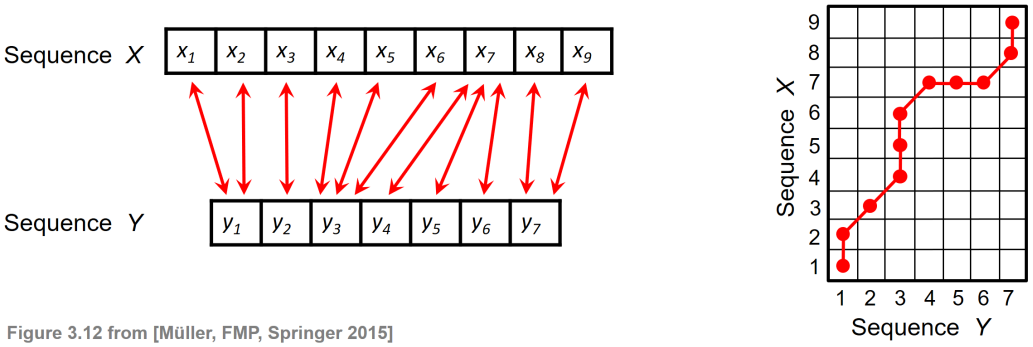


Figure 3.7: *Visualisation of DTW alignment method. Figure borrowed from [2].*

The algorithm starts by initialising the borders of the matrix and then fills in the rest based on the recursive relation that considers the cost of aligning two elements and the minimum cost of arriving at that alignment. The process is calculated involving dynamic programming to ensure that the optimal path is found. That is the one that minimally distorts the time axis of either sequence.

In sports analytics, DTW method proves to be crucial element used to precisely compare athletes' movements. This is especially applicable when compared activities vary slightly in timing or speed due to individual styles or environmental conditions. By detecting and accounting for these variations, DTW ensures meaningful comparisons that properly match athletes' performance characteristics. As depicted in Figure 3.8, DTW provides a more accurate alignment than Euclidean matching, which compares sequences point-by-point without adjusting for timing differences, thus respecting the natural flow of activity being analysed.

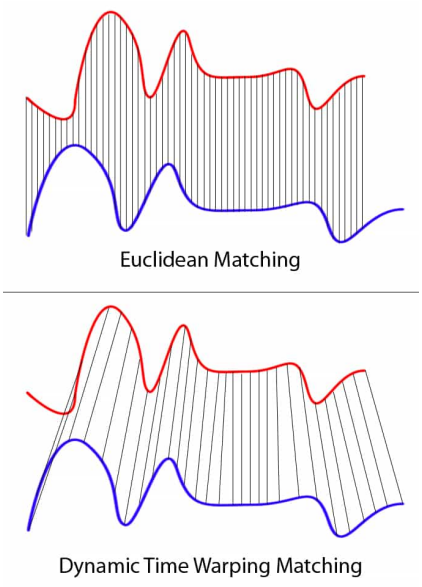


Figure 3.8: *Comparison of matching time series using euclidean and DTW methods. Note how highest point in red line is almost matched to the lowest point in blue line using euclidean matching, while using DTW the highest points are matched correctly. Figure borrowed from [3].*

3.5 Cosine Similarity

Cosine similarity is a metric used to measure the similarity between two vectors. It calculates the cosine of the angle between the vectors, providing a measure that ranges from -1 to 1. The cosine similarity is defined by the following formula:

$$\text{cosine\_similarity} = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{3.16}$$

where $\mathbf{x} \cdot \mathbf{y}$ is the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$, and $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are the magnitudes (or Euclidean norms) of the vectors. The result of this formula returns a value between -1 and 1, where 1 indicates that the vectors are identical, 0 indicates orthogonality (no similarity), and -1 indicates that the vectors are opposed.

In the context of measuring 2D human motion similarity, cosine similarity is useful because it standardises the output, ensuring a consistent approach

to similarity measurement. It has been therefore chosen by studies as the comparison method [1, 38, 39, 45].

The output ranges from -1 to 1 allows for clear interpretation of the similarity scores, making it easy to distinguish between similar and dissimilar motion sequences. This standardisation is important for reliable comparisons of motion embeddings, enabling the model to evaluate and distinguish between different motions based on their cosine similarity scores.

3.6 Summary

The final flow of the algorithm for analysing volleyball player movements is structured as follows:

1. **Preprocessing:** Input data is first adjusted for frame rate differences. Then normalisation algorithm is applied so that keypoints values are consistent across different videos. This part utilises both mentioned preprocessing steps.

2. **Phase Number Determination:** Based on scaled and normalised data, various possible phase configurations are calculated and presented to the user. User then chooses appropriate number of phases, thus injecting the expert knowledge into the system. This phase is completed using KernelCPD method.

3. **Phase Computation:** The optimal number of phases - given by user - is applied to both reference and compared video. Both movements are segmented by KernelCPD algorithm, ensuring consistent number of phases.

4. **Embedding Creation:** Body Part Embedding model is used to encode sequences of joints from each detected phase, to output representations for each body part and each phase. This part ensures following comparisons are conducted on reliable data.

5. **Alignment:** Dynamic Time Warping is used to align the embeddings. This process adjusts for temporal differences in both reference and compared movements. Because of this step, differences in speed or timings are not problematic while calculating the similarity scores.

6. **Similarity Calculation:** Similarities between the aligned embeddings are computed. The comparison is done using cosine similarities, which are computed using matching pairs in the DTW path between the embeddings. This score represents how close two movements are, with 1 being the same and -1 being totally opposite.

Structured approach described in this chapter ensures an accurate and reliable analysis of volleyball player movements. These steps ultimately lead to a more robust and effective methodology for analysing volleyball player movements. Following chapter will present results obtained from application of the system to volleyball keypoints data.

Chapter 4

Evaluation and Results

This chapter presents the test results of the system presented in the previous chapter. The chapter begins with a description of the data collection process, explaining how the data was collected, the structure of the dataset, and instructions for collecting similar data. Next, the validation approach is discussed, highlighting the subjective nature of movement similarity evaluation. This section explains how this task is unsupervised and why typical metrics are challenging to apply. The ideas for approach to evaluation are outlined. Following, the results of the phase detection algorithm are presented. The chapter then moves on to the results of the overall system, discussing how the final outputs are presented and justifying this presentation method. It also provides detailed insights into the final output. Final section of the chapter discusses results.

4.1 Data Collection

The data collection process for this research involved creating a dataset of volleyball attack movements, which includes a total of 199 videos. Among these, one video serves as the attack reference video; provided by Sport Vlaanderen - a sports organization based in Flanders. The remaining videos are various attempts to replicate or compare against this reference attack. Each video

is supplemented by a corresponding file that contains the extracted keypoint data and additional features.

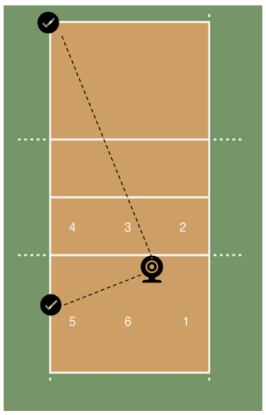


Figure 4.1: *Advised camera placement on the volleyball court.*

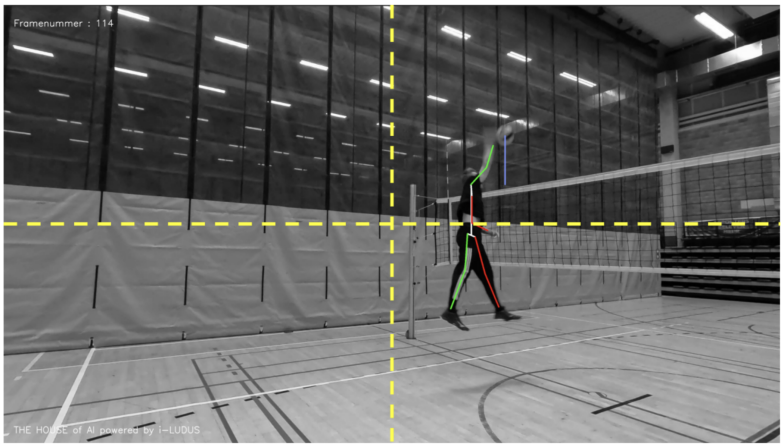


Figure 4.2: *Example of camera view, given advised placement presented in Fig. 4.1.*

The videos were collected in a controlled environment, following specific guidelines to ensure consistency, quality, and reproducibility. The videos were captured using either mobile phones or GoPro cameras, with each recording at 60 frames per second or higher to capture the fast-paced action with clarity (the only video with lower quality is the reference video at 24 FPS). The cameras were oriented in landscape mode to provide a wide-angle view. The placement of the camera was carefully considered to ensure optimal viewing angles and consistency across all videos. The cameras were positioned on flat ground and stabilised using either a tripod or a sturdy surface, ensuring minimal camera shake and stable footage. This setup was designed to

capture the full range of motion of the volleyball attack across the width of the image, while also ensuring that the full height of the movement was visible within the frame. To ensure clear and focused footage, several guidelines were followed regarding the content of the videos. The environment was controlled to avoid any extra people or balls in the background, eliminating potential distractions or obstructions that could interfere with the analysis. Resulting videos were therefore focused on the player and their movements, allowing for precise keypoint extraction and analysis. Figures 4.1 and 4.2 illustrate the optimal camera placement on the volleyball court, providing an overview of how the recordings were set up.

Each video is connected to a tabular file with extracted keypoints and additional features. The keypoint data contains the coordinates (x, y) of following points on the body: nose, both eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. In addition to the keypoint coordinates, the files also include several supplementary features provided by i-LUDUS. These features are the position of the ball, different joint angles, automatically annotated phases, handedness (indicating whether the hitter is left or right-handed), run-up steps, jump angle, position in relation to the ball, arm angle during ball hit, and video frame rate (FPS). While these additional features offer a comprehensive view of the volleyball attack movements, not all of them are directly relevant to the core focus of movement comparison. Nonetheless, these features enhance the richness of the dataset and provide valuable contextual information. Finally, the reference video is similarly connected with a tabular file, containing the same data as other files. It also serves as a ground truth, to which all other movements are compared.

This structured approach to data collection ensured high-quality footage and enabled a robust comparison between reference and training videos. The detailed guidelines for recording provided a solid foundation for the research, helping with comprehensive analysis of volleyball player movements.

4.2 Validation Approach

The validation approach in this study primarily relies on subjective assessments, recognising the detailed nature of evaluating volleyball player movements. The BPE [1] model, used as the core comparison algorithm, was evaluated in its original paper, and shown to be effective for comparing movements, which was mentioned in Chapter 2. However, the subjective nature of motion analysis and comparisons requires additional methods to assess the validity of this system.

One possible way to evaluate the phase detection by comparing them to automatic annotations, based on heuristics previously created by i-LUDUS. The first detected phase (run-up) starts with the beginning of video and ends while the feet are of the ground. Second phase (jump) starts while players feet are in the air and ends a frame before ball hit. Third phase (attack) is just few frames around the attack moment, detected by sound of ball hit. Fourth and final phase (landing) begins a few frames after attack moment and ends with the end of video. As the goal of this thesis is to create a data-driven unsupervised approach, these heuristics might not provide a good evaluation baseline as to the start and end of each phase. Nevertheless, the most important phase (attack) is helpful, as it must be clearly present within also the third phase of the automated annotation.

Another possibility involves visual inspection of the video and subjective assessment whether the segmented phases make sense. It incorporates expert knowledge into the analysis, using the insights of coaches or experienced players. To accomplish that, 30 randomly videos were chosen and manually inspected. It is also possible to compare different videos and see whether some specified points (e.g. breakpoints) correspond to similar poses. Furthermore, by applying Principal Component Analysis (PCA) to either x or y coordinates we can visually inspect, whether the breakpoints align with changes in movement representation. This is slightly more robust and scientific approach, yet it still misses out on a metric, as typically used in supervised learning. Because of that, the evaluation also tries to create a pseudo-metric, that can give us some intuition on how well the segmentation

part of system performs.

Furthermore, while comparing volleyball movements, it is important to note that just because two movements differ, it does not necessarily mean one is worse than the other. Different player techniques, changes in lines of approach or varying ball placement can affect the similarity of movements. For example, if player approaches an attack in parallel to the side line, and we want to compare it to the player that approaches in 45-degree angle to the side line, the camera view, and possibly entire motion will look different. This adds another layer of complexity in evaluation of the system. Moreover, different coaches might have differing ideas about how the ideal movement should look like, and even subtle differences can be problematic when assessing similarity. Therefore, it justifies a system that can adapt to the coach with the use of a reference video.

Finally, there is also the possibility of system evaluation by comparison of movements using hard-coded scores and rules. For example, such approach includes evaluating the elbow angle during the ball hit or lowest knee angle during jump moment. The biggest problem with this method is that it misses out on much of the information conveyed in the movement process. The hit moment represents only one frame, while the motion comparison score provides a more comprehensive assessment over several frames. That is why this approach will not be used in evaluation process in this thesis, as the focus is on movement; not frame evaluation.

4.3 Phase Detection Results

The following section presents results obtained from applying the phase detection algorithm to normalised keypoints data of volleyball movements. Different approaches outlined in previous section are described in detail. The section contains example plots of two randomly chosen movements and reference movement from dataset, presenting examples of how well the system performs. The plots are first described in detail and then discussed, arguing correctness of the performance by a qualitative analysis. Furthermore, a metric is presented in attempt to quantitatively assess algorithm's performance.

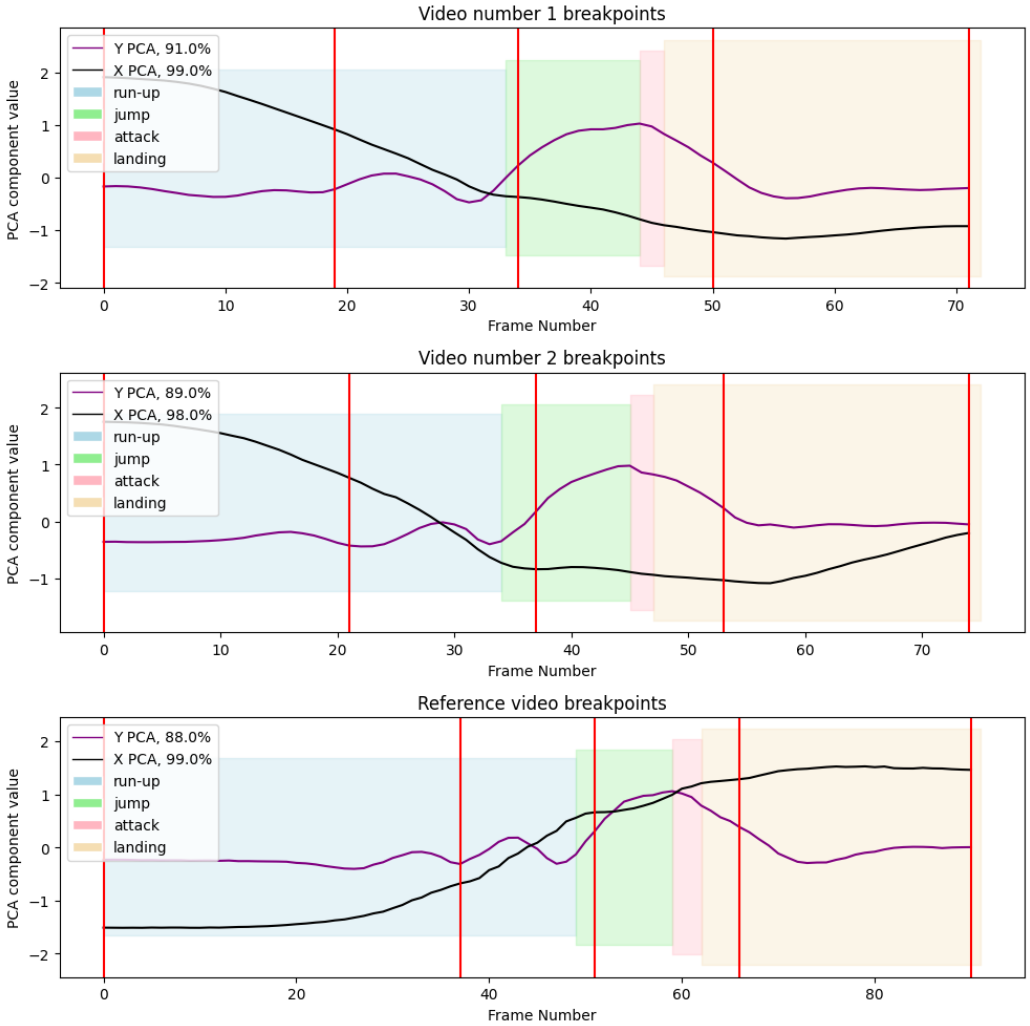


Figure 4.3: *Examples of detected breakpoints, with automatic phases as background colours. The purple and black lines represent PCA transformed values of 15 joints for x and y coordinates respectively. The labels also contain information on percentage of explained variance ratio by the given component. The vertical red lines segment the movement into detected phases. Leftmost and rightmost lines are artificially added to mark the beginning and end of movement. The 3 vertical lines between them are the breakpoints detected by algorithm described in previous chapter, thus segmenting each movement into 4 phases. The plot background is made with different colours, to represent phases detected by i-LUDUS algorithms. These are marked as 4 distinct phases: run-up (blue), jump (green), attack (red) and landing (beige). The X and Y axes of the plots correspond to frame number and value of given principal components.*

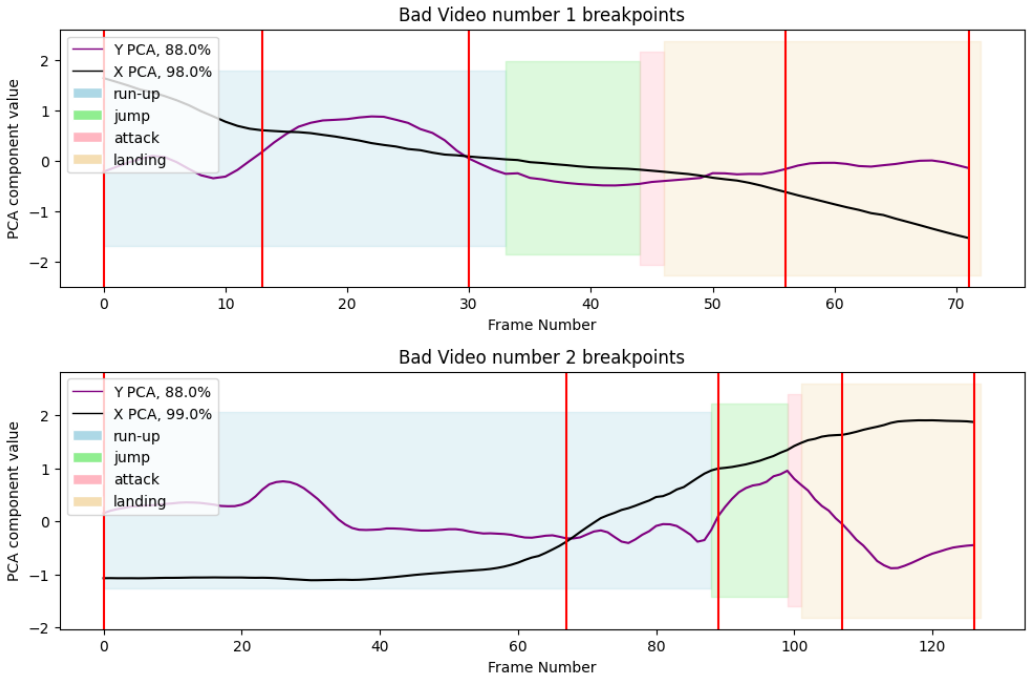


Figure 4.4: *Examples of incorrectly segmented videos. In first video the attack happens in phase 2. In second video the second phase, contains last 4 steps instead of 2, as in other analysed and inspected videos.*

Figure 4.3 depicts a visualisation of movements in 3 different videos. The following descriptions aim to help the reader with understanding the plots. Furthermore, the descriptions of phases are based on 30 randomly chosen videos, where in 29 of them there was consistent behaviour. Only data from two of these videos (Fig. 4.4) contain some inconsistencies.

The first phase in each plot corresponds to a standstill and first steps of movement - the least important part of entire movement. This can be confirmed by smallest changes in both X and Y coordinates. For example, in the reference video, there is almost no change in principal component X, in the first 20 frames. This correlates with the video, where during that time player does not move. We understand that this phase is also least important part of the movement, as it is far from attack frame and does not affect execution of further phases.

Second detected phase corresponds to the last two steps in the movement, including the start for jump. The first, upward bump, as seen in all video

plots corresponds to arms being thrown back an up; to build momentum for the jump (this is an important technique that allows players to jump higher). The second, downward bump, corresponds to arms and body coming down before the jump. This phase ends, approximately when players feet leave the ground.

We can observe, that in each video, during the 3rd phase there is a significant bump in value of Y component. Based on 30 manually inspected videos, it is apparent that start of this phase happens just as the feet of player leave the ground. Furthermore, we can confirm that in 29 out of 30 videos, this phase corresponds to the part of movement in which player makes a jump, hits the ball, and starts to fall. In all the plots, this is also confirmed by full overlap with attack phase, large overlap with jump phase and finally and small overlap with landing phase. The overlap is not exact with any of these phases, as they do not represent a true segmentation of movement. Also, the attack moment is detected as few frames around ball hit moment.

Finally, in the fourth and last phase, the player again comes to a standstill, which can be seen in all the presented plots as a graduate flattening of PCA lines. Although bumps corresponding to a wrist movement or attack do not necessarily look alike in all figures, the phase detection algorithm manages to correctly detect the breakpoints, making it a robust method for this problem.

The two inconsistent videos (out of 30), presented on Figure 4.4 depict some possible issues with the algorithm. In the first incorrectly segmented video, the movement starts later - while the player is doing last two steps before jump. This leads to jump and attack happening in second, instead of 3rd phase. Even though there is inconsistency here, it also shows that the heuristics not always work correctly, as they detected attack much later than it actually happened. In the second wrongly segmented video, the problem is with second phase, where the player does four steps instead of two.

Furthermore, Fig 4.5 presents plots similar to the ones in previous paragraph, but it also compares the breakpoints with skeleton poses. Each PCA and breakpoints plot is connected to corresponding plot representing the normalised position of player's body at each breakpoint detected by data driven phase detection algorithm (blue skeletons); positing during attack frame as

detected by heuristics (red skeleton).

It can be observed that the first skeleton, representing start of video is either during standstill (see video 1 and reference video) or at the start of movement (see video 2). Next is the skeleton representing start of phase with last two steps. Firstly, it can be seen that the displacement is much higher in second phase as compared to first phase. Secondly the pose is more dynamic as it leans forwards, indicating being in motion. The third skeleton, representing the start of jump looks similar in all presented samples, although in the video 1 the start of 3rd phase is detected a bit later. The skeleton has bent knees and arms in front, moving up, indicating start of the jump. The final blue skeleton presents breakpoint corresponding to end of jump. We can see that in all the plots this skeleton has its' feet higher than during on-ground movement. Furthermore, it can be seen that even the red skeletons - being the ones hitting the ball - differ substantially, highlighting the complex nature of movement analysis.

To determine the consistency of the phase detection algorithm, histograms in Fig. 4.6 were constructed. They present a distribution of each phases breakpoint; by frame number (Fig. 4.6a); and in relation to annotated attack frame (Fig. 4.6b). Red color is for end of first phase, green for end of second and blue for end of third. On both figures, histograms have 3 peaks, which correspond to start of phases two, three and four respectively. Fig. 4.6a contains single values above frame 60/70 due to one video containing substantially longer standstill at the beginning. There definitely is a consistency in detected breakpoints values. For example, most of the breakpoints corresponding to start 3rd phase were spotted to be 12 to 8 frames before defined attack frame. There can also be seen a clear distinction in Fig. 4.6b, which nicely presents reliability of proposed approach.

Finally, as mentioned in section 4.2, a pseudo-metric was developed to quantitatively assess phase detection algorithm's performance. The metric is calculated for entire dataset, by calculating the percentage of videos, where attack frame is inside third detected phase:

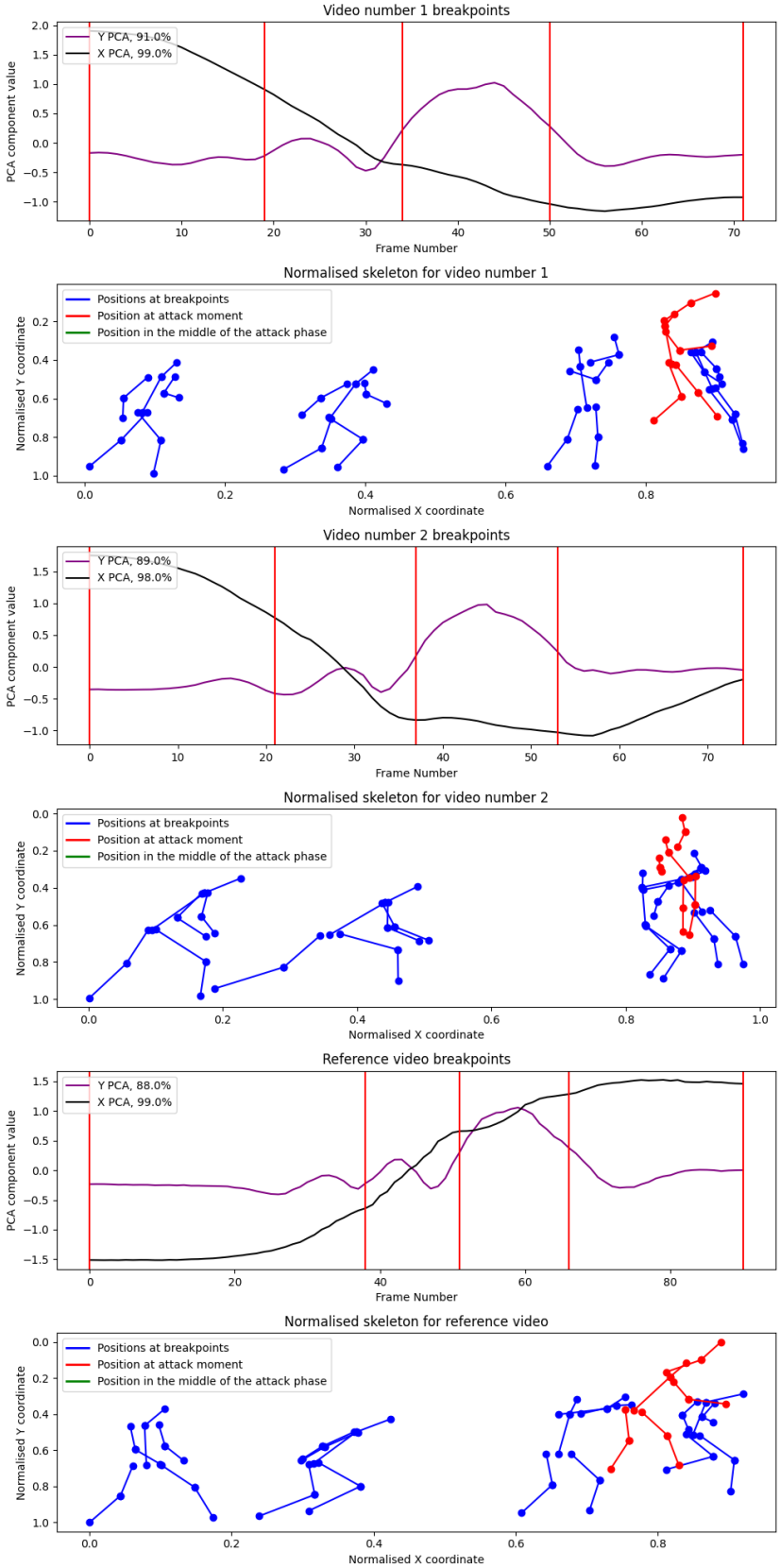


Figure 4.5: *Examples of detected breakpoints, including PCA transformed x and y coordinates, with corresponding skeletons visualisation.*

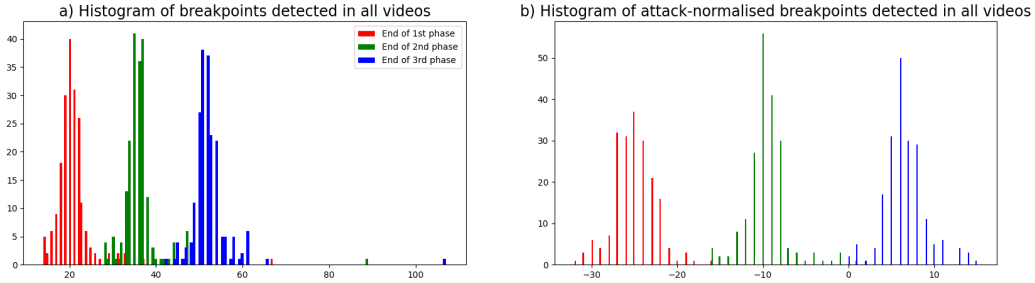


Figure 4.6: *Histograms of detected breakpoints. Red, green, and blue lines represent breakpoints distribution at end of 1st, 2nd and 3rd phase respectively. Figure a) shows how many breakpoints were detected at each phase (after fps aggregation); b) shows breakpoints, normalised based on attack frame.*

$$\left(\frac{\sum_{i=1}^{N} A_i}{N} \right) \times 100\% \tag{4.1}$$

where A_i represents the presence of an attack in the 3rd phase for the i-th file, and N is the total number of files. The resulting value was 98%, signifying presence of attack in the third phase almost every time.

In summary, the qualitative analysis provided in this section demonstrates that the phase detection algorithm is robust and effective for the available data. Detailed descriptions and discussions of the example plots, along with the presented pseudo-metric, confirm the algorithm's performance. This analysis offers the necessary background and results to validate the algorithm's capability in detecting phases within volleyball movements.

4.4 System Outputs

This section presents final outputs obtained from the system. These outputs apply simple transformations to the input motion video, to visually present comparison result to the user.

Figure 4.7 visualises the system's output across different frames; selected based on detected keypoints. Top left corner of the video displays information on similarity of each body part during given phase. The current phase is highlighted in green in the top right corner, while other phases are displayed

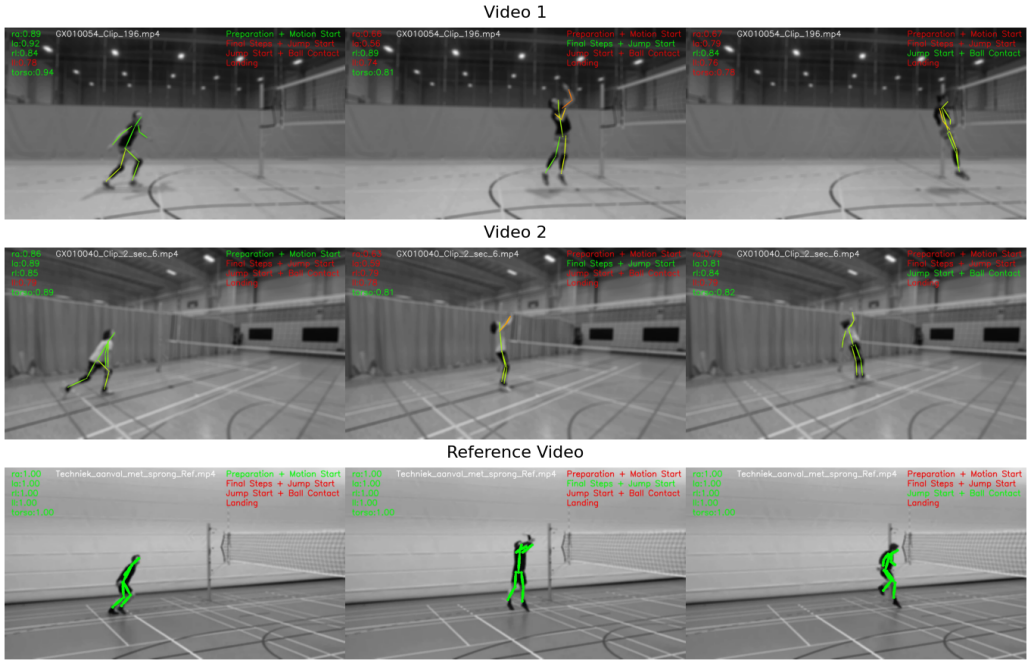


Figure 4.7: *Breakpoint frames from each of previously selected video, including comparison scores and phase annotations.*

in red. The upper middle part of the output contains the name of the video file on which the analysis was conducted. The skeletons are plotted onto the body of person that performs the movement. Colours of each skeleton body part are chosen based on colour palette that diverges from red to green as presented on Fig. 4.8. The minimum score on the palette was obtained by comparing all video files in the dataset to the reference movement and selecting the minimum similarity value. The maximum value is 1 - the maximum value of cosine similarity.

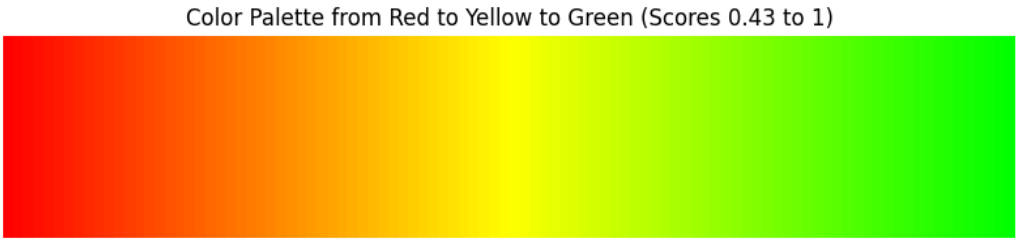


Figure 4.8: *Color palette used to highlight differing body parts, where red represents dissimilarity and green - similarity.*

For reference video in all the phases the similarity score is equal to one for all

body parts and all the body parts are green. This signals perfect similarity of two movements - which is expected, since it is the same movement being compared. Given example also serves as a sanity check for the entire system, demonstrating its reliability. In the case of first two presented videos, it is hard to judge, whether the similarity scores are correct. Based on single frames it is impossible to assess the system performance, therefore these examples serve purely demonstration purposes. As mentioned in Chapter 2, the Body Part Embedding model used for motion similarity assessment has been thoroughly evaluated by the authors. The BPE model's ability to capture fine-grained motion details justifies its use and enhances the system's overall effectiveness.

Chapter 5

Conclusion

This thesis developed and evaluated an AI-based system for analysis and comparison volleyball player movements. The primary aim was to address the limitations in current volleyball analysis methods by providing a more objective, precise, automated, and data-driven system for movement analysis. By incorporating advanced preprocessing techniques, phase detection algorithm, and a deep learning model, this research has contributed significantly to the field of sports science - particularly in volleyball training.

The phase detection algorithm's performance, as discussed in the Chapter 4, demonstrated effectiveness in segmenting volleyball movements into meaningful phases. Upon verification, the detected phases can be defined as: start of movement; last two steps before jump; jump and attack; landing. The qualitative analysis, including manual inspection of 30 videos, showed accurate phase identification, with minimal changes in PCA components during the movement start phase; significant body displacement during the last two steps; a marked bump in the Y component during the jump and attack; and gradual flattening of PCA lines during the landing. Skeleton visualisations further validated these findings, depicting the expected actions correctly. Furthermore, quantitative analysis, supported by a pseudo-metric with a 98% consistency rate of attack frames within the third phase, supports claims to the algorithm's reliability.

This thesis has therefore made significant effort in improving volleyball train-

ing methodologies through the application of AI and pose estimation techniques. The developed system not only provides a valuable tool for performance evaluation and improvement, but also lays the groundwork for future innovations in sports analytics.

5.1 Contributions

The system's phased comparison approach, utilising key joint position extraction, phase detection, and BPE [1], has proven to be effective in segmenting and analysing complex volleyball movements. The results indicate that the system can reliably identify and compare different phases of a volleyball attack, offering detailed feedback that can be invaluable for coaches and athletes. This capability addresses the need for more granular and objective analysis in sports training, moving beyond traditional subjective assessments.

One of the main achievements of this research is the integration of expert knowledge into the phase detection process, which ensures that the system's outputs are aligned with practical performance insights. The improvement of use of DTW for temporal alignment of embeddings, further improves the system outputs, allowing it to handle variations in movement speed and timing effectively.

Furthermore, this research also contributes to the field of motion analysis, by using FPS aggregation and bounding box normalisation techniques. These methods help with efficient and reliable comparison of movements originating from videos made with different cameras and of different body sizes.

5.2 Limitations

The system's reliability in assessing movement similarity was partly confirmed by achieving perfect similarity scores for all body parts using a reference video, demonstrating accuracy in comparing identical motions. However, as discussed in the Chapter 4, the subjective nature of evaluating movement similarity and the limitations of automated annotations, show the com-

plexity of this task. Comparing motion similarity scores to scores given for only one frame is insufficient for comprehensive evaluation.

A major limitation is the lack of evaluation for phase detection algorithm, using fully hand-annotated videos, which would provide a more reliable ground truth. Additionally, the system performance has not been thoroughly validated with feedback from professional coaches, which is important factor, that would ensure applicability in real-world scenarios.

In conclusion, while we know that the BPE similarity outputs have been evaluated (as shown in Chapter 2), the full system developed in this thesis lacks means of evaluation. That is why future improvements should focus on collecting better data, expanding datasets, and collecting feedback from professional coaches. These efforts should confirm system's robustness in comparing movements or provide necessary feedback for improvement.

5.3 Future Work

To improve the phase detection algorithm, future work should involve using fully hand-annotated videos, providing a more reliable ground truth for validation. Additionally, exploring the application of this algorithm to other sports or physical activities could demonstrate its versatility and scalability. Furthermore, improvements of phase detection algorithm, such as intelligently connecting corresponding phases, can be made to deal with problems shown in Chapter 4.

For labelled data, integrating advanced classification algorithms like Long Short-Term Memory (LSTM) networks could improve phase detection and classification accuracy. These efforts can potentially improve algorithm's effectiveness and make it widely applicable in motion analysis field. With use of labelled data based on biomechanics studies, algorithms could be implemented to provide objective and scientifically based movement segmentation.

Moreover, expanding the dataset by incorporating feedback from professional coaches is crucial for improving the system's accuracy and practical applications. Such efforts will improve the system's reliability in analysing volleyball

and potentially other sport movements, expanding the applications of proposed system.

Finally, implementation of solutions based on 3D pose data is another promising direction, as detailed three-dimensional insights into athletes' movements could significantly improve performance optimisation and injury prevention strategies.

Bibliography

[1] Jonghyuk Park, Sukhyun Cho, Dongwoo Kim, Oleksandr Bailo, Heewoong Park, Sanghoon Hong, and Jonghun Park. A body part embedding model with datasets for measuring 2d human motion similarity. *IEEE Access*, 9:36547–36558, 2021.

[2] Meinard Müller. *Fundamentals of music processing: Audio, analysis, algorithms, applications*, volume 5. Springer, 2015.

[3] WikimediaCommons. Euclidean vs dtw.

[4] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective. *WIREs Data Mining and Knowledge Discovery*, 13(5):e1496, 2023.

[5] Aritz Badiola-Bengoa and Amaia Mendez-Zorrilla. A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. *Sensors*, 21(18), 2021.

[6] João Claudino, Daniel Capanema, Thiago Souza, Julio Serrao, Adriano Pereira, and George Nassis. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Medicine - Open*, 5, 07 2019.

[7] Haoming Chen, Runyang Feng, Sifan Wu, Hao Xu, Fengcheng Zhou, and Zhenguang Liu. 2d human pose estimation: a survey. *Multimedia Systems*, 29(5):3115–3138, Oct 2023.

[8] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. Deep learning-based

human pose estimation: A survey. *ACM Comput. Surv.*, 56(1), aug 2023.

[9] Yiwen Gu, Shreya Pandit, Elham Saraee, Timothy Nordahl, Terry Ellis, and Margrit Betke. Home-based physical therapy with an interactive computer vision system. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2619–2628, 2019.

[10] Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu. Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 374–382, New York, NY, USA, 2019. Association for Computing Machinery.

[11] Omar Tarek, Omar Magdy, and Ayman Atia. Yoga trainer for beginners via machine learning. In *2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, pages 75–78, 2021.

[12] Lei Yang, Yingxiang Li, Degui Zeng, and Dong Wang. Human exercise posture analysis based on pose estimation. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, volume 5, pages 1715–1719, 2021.

[13] Steven Chen and Richard R. Yang. Pose trainer: Correcting exercise posture using pose estimation, 2020.

[14] Ashish A Keoliya, Swapnil U Ramteke, Manali A Boob, and Kamya J Somaiya. Enhancing volleyball athlete performance: A comprehensive review of training interventions and their impact on agility, explosive power, and strength. *Cureus*, 16(1), January 2024.

[15] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[16] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields, 2019.

[17] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time, 2022.

[18] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, Peter Gehler, and Bernt Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation, 2016.

[19] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation, 2019.

[20] Debapriya Maji, Soyeb Nagori, Manu Mathew, and Deepak Poddar. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss, 2022.

[21] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019. Accessed: 2024-04-25.

[22] Denis Tome, Chris Russell, and Lourdes Agapito. Lifting from the deep: Convolutional 3d pose estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[23] L. Ponce Cuspinera, Sakura Uetsuji, F. Morales, and Daniel Roggen. Beach volleyball serve type recognition. pages 44–45, 09 2016.

[24] Fasih Haider, Fahim Salim, Vahid Naghashi, Sena Busra Yengec Tasdemir, Izem Tengiz, Kubra Cengiz, Dees Postma, Robby van Delden, Dennis Reidsma, Bert-Jan van Beijnum, and Saturnino Luz. Evaluation of dominant and non-dominant hand movements for volleyball action modelling. In *Adjunct of the 2019 International Conference on Multimodal Interaction*, ICMI '19, New York, NY, USA, 2019. Association for Computing Machinery.

[25] Thomas Kautz, Benjamin H. Groh, Julius Hannink, Ulf Jensen, Holger Strubberg, and Bjoern M. Eskofier. Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery*, 31(6):1678–1705, Nov 2017.

[26] Yufan Wang, Yuliang Zhao, Rosa H. M. Chan, and Wen J. Li. Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist. *IEEE Access*, 6:13758–13765, 2018.

[27] Evan Quinn and Niall Corcoran. The automation of computer vision applications for real-time combat sports video analysis. In *European Conference on the Impact of Artificial Intelligence and Robotics*, volume 4, pages 162–171, 2022.

[28] Chien-Chang Chen, Chen Chang, Cheng-Shian Lin, Chien-Hua Chen, and I. Cheng Chen. Video based basketball shooting prediction and pose suggestion system. *Multimedia Tools and Applications*, 82(18):27551–27570, Jul 2023.

[29] Yung-Che Li, Ching-Tang Chang, Chin-Chang Cheng, and Yu-Len Huang. Baseball swing pose estimation using openpose. In *2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*, pages 6–9, 2021.

[30] Jen Jui Liu, Jacob Newman, and Dah-Jye Lee. Body motion analysis for golf swing evaluation. In George Bebis, Zhaozheng Yin, Edward Kim, Jan Bender, Kartic Subr, Bum Chul Kwon, Jian Zhao, Denis Kalkofen, and George Baciu, editors, *Advances in Visual Computing*, pages 566–577, Cham, 2020. Springer International Publishing.

[31] Jen-Jui Liu, Jacob Newman, and Dah-Jye Lee. Using artificial intelligence to provide visual feedback for golf swing training. *Electronic Imaging*, 33(6):321–1–321–1, 2021.

[32] Chen-Chieh Liao, Dong-Hyun Hwang, and Hideki Koike. Ai golf: Golf swing analysis tool for self-training. *IEEE Access*, 10:106286–106295, 2022.

[33] Lijin Zhu. Computer vision-driven evaluation system for assisted decision-making in sports training. *Wireless Communications and Mobile Computing*, 2021:1865538, Aug 2021.

[34] Amit Nagarkoti, Revant Teotia, Amith K. Mahale, and Pankaj K. Das. Realtime indoor workout analysis using machine learning computer

vision. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1440–1443, 2019.

[35] Elham Saraee, Saurabh Singh, Kathryn Hendron, Mingxin Zheng, Ajjen Joshi, Terry Ellis, and Margrit Betke. Exercisecheck: Remote monitoring and evaluation platform for home based physical therapy. pages 87–90, 06 2017.

[36] Atima Tharatipyakul, Kenny T. W. Choo, and Simon T. Perrault. Pose estimation for facilitating movement learning from online videos. *CoRR*, abs/2004.03209, 2020.

[37] Jiayao Emily Li and Haridhar Pulivarthy. Balletnettrainer: An automatic correctional feedback instructor for ballet via feature angle extraction and machine learning techniques. In *Proceedings of the International Conference on Industrial Engineering and Operations Management, ARQuest SSERN International, Kirkland, WA, United States*, pages 603–613, 2021.

[38] Hobeom Jeon, Dohyung Kim, and Jaehong Kim. Human motion assessment on mobile devices. In *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1655–1658, 2021.

[39] Jiangkun Zhou, Wei Feng, Qujiang Lei, Xianyong Liu, Qiubo Zhong, Yuhe Wang, Jintao Jin, Guangchao Gui, and Weijun Wang. Skeleton-based human keypoints detection and action similarity assessment for fitness assistance. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, pages 304–310, 2021.

[40] Adobe Systems. Mixamo. `https://www.mixamo.com/#/`.

[41] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. Ntu rgb+d 120: A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2684–2701, October 2020.

[42] Charles Truong, Laurent Oudre, and Nicolas Vayatis. ruptures: change point detection in python, 2018.

[43] Alain Celisse, Guillemette Marot, Morgane Pierre-Jean, and Guillem Rigaill. New efficient algorithms for multiple change-point detection with kernels, 2017.

[44] Kfir Aberman, Rundi Wu, Dani Lischinski, Baoquan Chen, and Daniel Cohen-Or. Learning character-agnostic motion for motion retargeting in 2d. *ACM Transactions on Graphics*, 38(4):1–14, July 2019.

[45] Ega Hegarini, Achmad Benny Mutiara, Adang Suhendra, Mohammad Iqbal, and Bheta Agus Wardijono. Similarity analysis of motion based on motion capture technology. In *2016 International Conference on Informatics and Computing (ICIC)*, pages 389–393, 2016.

[46] Trevor Lynn. Pose estimation algorithms: History and evolution. `https://blog.roboflow.com/pose-estimation-algorithms-history/`, 2023. Accessed: 2024-04-25.

[47] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, October 2012.