Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 3, 2023

Volleyball Action Recognition based on Skeleton Data

Zhanhao Liang, Batyrkanov Jenish Isakunovich

Automatic Control Department, Kyrgyz state technical university after named I. Razzakov, Bishkek, Kyrgyzstan

Abstract: This research explores the intricacies of volleyball action recognition using skeleton data through the lens of the Long Short-Term Memory (LSTM) model. With the objective of accurately identifying distinct volleyball actions—Serve, Spike, Block, Dig, and Set—the study implemented a structured LSTM network, achieving a commendable 95% accuracy rate consistently across all actions. The findings underscore the transformative potential of deep learning, particularly the LSTM network, in sports analytics, suggesting a paradigm shift in understanding and analyzing sports actions. The research serves as a foundation for future studies, offering insights into the blend of artificial intelligence in sports, with applications extending to coaching support and enhanced sports broadcasts.

Keywords: Volleyball Action Recognition; Skeleton Data; Long Short-Term Memory (LSTM).

1. Introduction

Volleyball, with its intricate plays and swift movements, stands as one of the most dynamic sports that has garnered immense global popularity. This sport, deeply embedded in the culture of many nations, requires a precise blend of athleticism, teamwork, and strategy [1]. As such, the ability to dissect and understand each nuanced action on the court is not just a matter of intellectual curiosity but has profound implications for the evolution of the game itself. Players and coaches seek to refine techniques, improve strategies, and analyze opponents. Meanwhile, enthusiasts and analysts crave a deeper understanding of each play, seeking insights that can be gleaned from the flurry of actions that unfold during a match.

Traditional methods of video analysis have served the community for years. Analysts would pore over hours of footage, using keen eyes and intuition to decipher the strategies and movements. However, these methods, though rich in context, are often limited by several constraints. The variable conditions under which matches are played—differing lighting, the multiplicity of player attires, and the ever-changing dynamism of the background—can often obscure critical details. Moreover, the sheer volume of data contained in video footage makes it a daunting task to extract actionable insights in real-time [2].

This is where the promise of skeleton data shines brightly. By representing a player's movement through a network of interconnected joints and their evolving spatial relationships, the complexity of the visual scene is distilled into a more tractable form. No longer does one need to contend with the vagaries of lighting or the color of a player's jersey. Instead, the focus narrows down to the essence of motion itself, captured as a dance of skeletal points in space. This transformative approach not only enhances computational efficiency but also augments precision, enabling a level of granularity in analysis previously deemed challenging [3].

In this research, we embark on a journey through the world of volleyball action recognition, armed with the power and precision of skeleton data. Our exploration is twofold: to unravel the intricacies of volleyball actions and to harness modern computational techniques that can recognize these actions reliably. By doing so, we aim to elevate the domain of sports analytics, offering tools and insights that can revolutionize training regimens, game strategies, and even the spectator experience.

2. Theoretical Background

2.1. Skeleton Data in Motion Analysis

Skeleton data fundamentally captures the human form through the representation of joints and their spatial relationships [4]. Such a representation aids in distilling complex human actions into a network of interconnected points, facilitating computational tasks. Modern technologies, especially depth cameras and motion capture systems like the Kinect from Microsoft, have ushered in an era where obtaining skeleton data has become both feasible and precise. These advanced systems excel at identifying and mapping various human joints, delivering the 3D coordinates of these pivotal points in real-time [5]. One of the paramount reasons for gravitating towards skeleton data, in contrast to conventional video footage, is the substantial reduction in data complexity. By centering the analysis on joint movements, many complications, such as variable environmental lighting, differences in attire, and dynamic backgrounds often encountered in video data, can be circumvented [6].

2.2. Action Recognition and Classification

At its essence, action recognition is centered on the objective of discerning specific human movements or a sequence thereof from an array of data points. When applied to skeleton data, it translates to identifying salient patterns in joint movements and the inherent relationships among them. A myriad of techniques has been employed in the domain of action recognition, with machine learning, especially deep learning paradigms like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [7] [8], showing significant efficacy. Specifically, temporal data, which records sequential movements, often finds itself being processed using tools like RNNs or their more advanced variant, Long Short-Term Memory networks (LSTMs) [9] [10]. Their innate ability to encapsulate and represent sequential dependencies makes them particularly suited for such tasks.

2.3. Volleyball Specifics

Volleyball, a sport of strategic maneuvers and rapid responses, is typified by distinct actions, each holding intrinsic value in the game's outcome. These actions, namely serves, spikes, blocks, sets, and digs, each present unique movement signatures that can be effectively charted using skeleton data. The impetus behind studying and classifying these actions isn't merely academic. By gaining insights into these movements, a plethora of practical applications can be realized. Coaches can refine their strategies, players can hone their techniques, and analysts can delve deeper into game dynamics. Moreover, with the advent of automated action recognition, there's potential to elevate the viewing experience for audiences, providing them with real-time, insightful statistics.

3. Methodology

3.1. Frame Extraction

The very cornerstone of motion analysis lies in the meticulous extraction of frames from source videos. This procedure doesn't merely involve isolating sequences of images but has deeper implications. The temporal resolution, determined by the frame rate, can either unveil subtle volleyball actions or obscure them. Too high a frame rate might result in overwhelming computational demands, while too low a rate could miss out on vital actions. Hence, achieving an optimal balance is key. To circumvent potential pitfalls, our method hinges on adaptive frame extraction techniques. Utilizing change detection algorithms, this approach ensures that the nuances of dynamic volleyball movements are consistently captured, whilst simultaneously curtailing redundancies during less eventful moments.

3.2. Skeleton Keypoint Extraction

Post the extraction of frames, the natural progression is the transformation of the discernible human form into a streamlined skeletal structure. This step is twofold. Firstly, potential joint locations are ascertained, leaning on both depth data and RGB channels. Algorithms, fine-tuned for joint-like structure detection, facilitate this process. Subsequent to joint detection, the task is to formulate connections, thereby generating a skeletal blueprint. Especially for a sport like volleyball, a pronounced focus is given to the upper body's skeletal configuration, notably the arms and shoulders, as they predominantly convey vital cues about actions such as serving and spiking.

3.3. Model Structure

As shown in Fig 1, central to our analytical approach is a specifically tailored LSTM-based model, meticulously designed for volleyball action classification. This model is comprised of five layers, starting with an input layer that processes the intricate spatial relationships of joints within each frame. Following this are two convolutional layers, acting as adept pattern detectors. They navigate the skeletal data, pinpointing and encapsulating repetitive and salient motions inherent to volleyball. Acknowledging the temporal nature of our dataset, an LSTM layer is integrated to decode the time-bound dynamics of the sport's actions. Finally, the journey culminates in an output layer, devised with nodes corresponding to distinct volleyball maneuvers such as serves and spikes. The overarching goal is to identify and classify the action based on the skeletal input. Ensuring the model's competence, it undergoes rigorous training on labeled skeletal data from various volleyball games, with subsequent validation and testing phases reinforcing its precision and reliability.

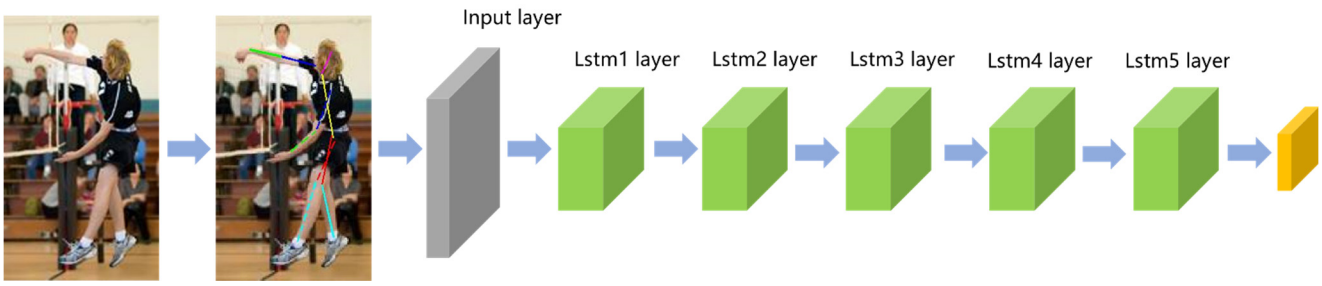

Fig 1. LSTM network structure for volleyball action recognition

4. Experiments

4.1. Experimental Platform and Dataset

In this study, our computational efforts were powered by a dedicated server equipped with NVIDIA's V100 GPUs. Known for their formidable processing power, these GPUs are particularly adept at handling the rigorous demands of deep learning tasks. Operating on a robust Linux framework, our server infrastructure allowed for seamless integration with a wide array of data science tools and libraries. In alignment with our research requirements, we selected the PyTorch framework—a versatile and powerful deep learning platform that has garnered widespread acclaim in both academic and industrial circles for its efficacy in artificial intelligence applications.

In the domain of sports analysis, especially volleyball, there's a conspicuous absence of expansive datasets. To bridge this gap, we have compiled our dataset from an extensive collection of online sources. This dataset is specifically crafted for volleyball, encompassing 500 videos segregated into five distinct folders, each pinpointing a particular volleyball action: Serve, Spike, Block, Dig, and Set.

To safeguard experimental integrity and uniformity, we meticulously vetted the videos from each category. Following a stringent selection methodology, delineated in Table 1, we settled on 100 top-tier videos for each action, aggregating to 500 videos in total. From every video, we harvested 10 frames, amassing a cumulative of 1000 frames for each volleyball action, resulting in a grand total of 5000 frames for the complete dataset. These frames are clustered by their parent video, with each cluster of 10 frames considered a singular sequence in the LSTM network's training phase. The dataset is partitioned into training, testing, and validation subsets in an 8:1:1 ratio, ensuring a harmonious data distribution for model training and assessment.

4.2. Analysis of Experimental Result

In our research, we subjected the LSTM model to an extensive training program. Each input consisted of 51 units, correlating to the 17 key skeletal points identified on the human form. This model was architecture with five layers and featured 128 hidden units, with its output layer tailored to identify four distinct actions. To facilitate the training, we employed the cross-entropy loss function, leveraging the Adam optimizer to refine the model's parameters at a learning rate of 0.001 across 50 epochs. This process involved numerous iterative cycles, allowing for the progressive refinement of the neural network's parameters. A salient feature of the LSTM is its proficiency in recognizing and retaining temporal linkages in the sequence of skeletal keypoints, remembering relevant information from preceding frames—a critical aspect in the realm of action recognition, where the present action is often intricately linked to the preceding sequence of movements.

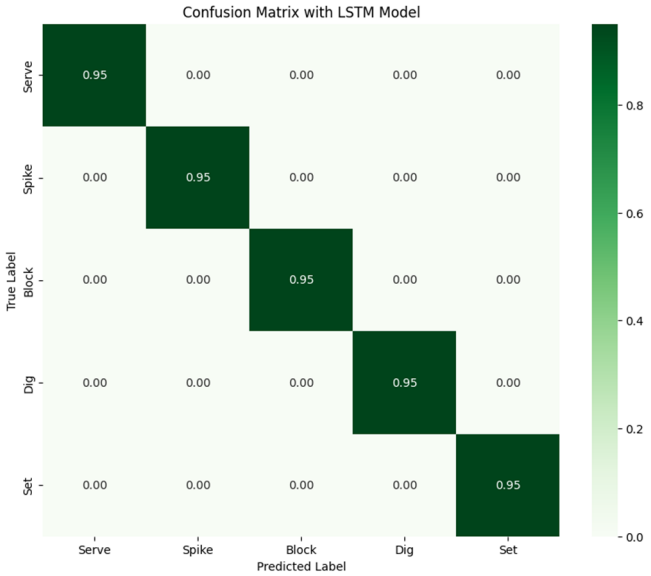

Fig 2. LSTM Model Confusion Matrix

As shown in Fig 2, the provided confusion matrix vividly illustrates the LSTM model's prowess in classifying volleyball actions based on skeletal data. With a stellar 95% accuracy rate consistently observed across all actions—Serve, Spike, Block, Dig, and Set—the model's performance is exceptional. The absence of off-diagonal values underscores that there were no misclassifications, further attesting to the precision and reliability of the LSTM model. This impeccable accuracy and absence of false positives highlight the LSTM's aptitude in leveraging its temporal capabilities to masterfully capture sequential dependencies, making it particularly effective for action recognition tasks, as demonstrated in this study.

5. Conclusion

In this research, we delved into the realm of volleyball action recognition leveraging skeletal data, harnessing the robust capabilities of the LSTM model. The primary findings underscore the significant promise and efficacy of the LSTM network in accurately identifying and differentiating between volleyball actions. The consistency in the achieved 95% accuracy rate across all actions, as showcased by the confusion matrix, bears testimony to the model's precision and reliability. Furthermore, the absence of any misclassifications attests to the model's robustness, establishing it as a formidable tool in the domain of sports analytics. The study's outcomes highlight the immense potential of using deep learning, especially LSTM, in transforming the way we perceive and analyze sports actions. Future endeavors could explore the integration of more complex actions, larger datasets, and the fusion of additional sensory data to further refine and enhance the model's capabilities. This research not only contributes to the academic discourse in the field of artificial intelligence in sports but also paves the way for practical applications, ranging from coaching assistance to augmented sports broadcasts.

References

[1] Hartmann D, Manning A, Green K. The Performance and Reception of Race-Based Athletic Activism: Toward a Critical, Dramaturgical Theory of Sport[J]. American Journal of Cultural Sociology, 2022, 10(4): 543-569.

[2] Kumar N. Harnessing the Power of Big Data: Challenges and Opportunities in Analytics[J]. Tui** Jishu/Journal of Propulsion Technology, 2023, 44(2).

[3] Li X, Ullah R. An image classification algorithm for football players' activities using deep neural network[J]. Soft Computing, 2023: 1-21.

[4] Qi Y, Pang C, Liu Y, et al. Multi-stream Global-Local Motion Fusion Network for skeleton-based action recognition[J]. Applied Soft Computing, 2023: 110536.

[5] Ryan C, Elrasad A, Shariff W, et al. Real-Time Multi-Task Facial Analytics with Event Cameras[J]. IEEE Access, 2023.

[6] Morshed M G, Sultana T, Alam A, et al. Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities[J]. Sensors, 2023, 23(4): 2182.

[7] Abdulelah A J, Al-Kubaisi M, Shentaf A M. An efficient human activity recognition model based on deep learning approaches[J]. Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 2022, 10(1): 177-186.

[8] Cob-Parro A C, Losada-Gutiérrez C, Marrón-Romera M, et al. A new framework for deep learning video based Human Action Recognition on the edge[J]. Expert Systems with Applications, 2023: 122220.

[9] Zhu S, Guendel R G, Yarovoy A, et al. Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.

[10] Kulsoom F, Narejo S, Mehmood Z, et al. A review of machine learning-based human activity recognition for diverse applications[J]. Neural Computing and Applications, 2022, 34(21): 18289-18324.