Athlete Action Recognition in Sports Video: A Survey

K. Kausalya and S. Kanaga Suba Raja

1 Introduction

The sports video analysis concentrates on the team's and individual performance of the player. With the help of a human vision system, the actions of the players are easily recognized and observed. But monitoring a player's action using human labor is highly expensive and also there is a chance of misrecognition of the action. So, there is a need for some automated way of recognizing the actions. The learning algorithms could build a machine that can apparently realize the action which will be needful for sports video action recognition. The learning algorithms like machine learning and deep learning are used to easily predict and recognize the data. This automatic system not only helps in improving the player's performance but also helps the coach to train the players: it is also able to analyze the game rules, like the movements of the players and team performances.

Considering the advantages of learning algorithms, this paper attempts to focus the survey on sports video action recognition. It is surveyed under two categories mainly: (i.e.) athlete tracking as well as sports video action recognition. First in athlete tracking the player's identity is to be predicted in [48]; the author differentiates the jersey color of the players using MAP detection and later separates the numbers and then with the help of a template matching the players' identities are found out. In [49] the author pro- poses a novel model that automatically identifies and tracks the players. In order to overcome issues like occlusions, players' ambiguous appearance and motion patterns outside the field, Chun-Wei lu proposed a novel framework that

K. Kausalya (✉)
Department of Information Technology, Easwari Engineering College, Chennai, India
e-mail: kausalyamurthy@gmail.com

S. Kanaga Suba Raja
Easwari Engineering College, Chennai, India

R. J. Kannan et al. (eds.), *International Virtual Conference on Industry 4.0*, Lecture Notes in Electrical Engineering 1003, https://doi.org/10.1007/978-981-19-9989-5_14

can achieve 82% and 79% for tracking algorithm and player identification. In [51] the author proposes a model to identify if a player is idle for some period of time without any action.

In sports video action recognition, the player's actions are recognized based on the trained datasets or on learning algorithms. In a few datasets like the BEAVOLL dataset [6], UCF Sports [10], Olympic [50], Hockey dataset [15], THETIS [16, 30], and HMDB [16, 31] are mainly used to detect the players' actions. In the above dataset, some of the actions like free hit, serve, dig, non-action, pass, spike, block, save, walk and run can be achieved. And learning algorithms like Convolution Neural Network (CNN) along with Recurrent Neural Network (RNN) be implemented to recognize the actions in sports videos. Each of the above phases is important and the purposes vary based on particular monitoring. All the above listed issues are to be solved that enhance the performance optimization problems using current emerging technologies. Figure 1 represented the work flow of the sports video action recognition. Where, in initial stage the input video is framed then its feature is extracted followed by detecting the objects. In Final stage, the motion tracking followed by action classification is done. In this flow, the action recognition of sports video is done generally.

The organization of the paper has seven sections: Sect. 2 of the paper represents the related works that include athlete tracking, sports video action recognition and datasets based on existing studies. Section 3 describes the analysis of the techniques.

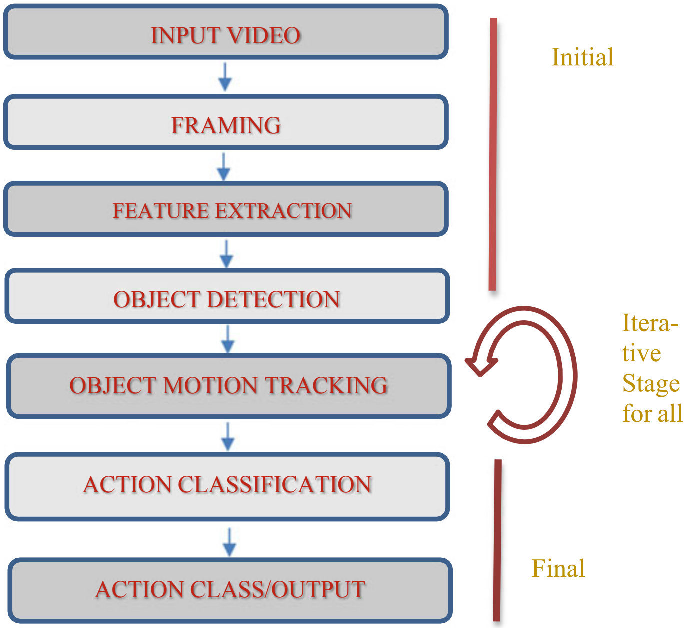


Fig. 1 Work flow of the sports video action recognition

Section 4 determines the inference made from the papers. Section 5 discusses the real world applications. Section 6 explains about the future direction of the proposed research with Sect. 7 concludes the paper.

2 Literature Review

Researchers in the area of pattern recognition and computer vision had a great interest in action recognition. When it comes to sports video processing, it is considered to be one of the most interesting and challenging topics. Frequent occlusion, background clutter, out of view, fast motion and motion blur are some of the action recognition issues in sports video analysis. The problems are investigated and categorized by other works. This section includes the studies based on athlete tracking followed by the action recognition in sports videos and concludes with the list of datasets that are used for training and testing purposes in sports videos.

2.1 Athlete Tracking

Athlete tracking is standard practice for monitoring the player's performance, behavior and also to reduce injury risk. It would be helpful for coaches to train the players by identifying their faults and train them accordingly [6]. Athlete tracking and action recognition are considered to be the two major problems that are associated with each other. In existing studies, they have been considered separately. Longteng Kong et al., proposed a system that solves the existing problem. The new dataset BeaVoll is released for analyzing the sports video. Here, both athlete tracking and action recognition are developed and work as a joint framework. A Scaling and occlusion robust tracker (SORT) based on Compressive Tracking (CT) is proposed for the tracking module. Long-term Recurrent Region-guided Convolutional Network takes the tracking results as input for action recognition.

Features of different sizes are extracted by the assigned SPP-net along with the Long and short-term memory network that is been adapted to model action dynamics. Figure 2 shows the screenshots of the BeaVoll datasets tracking results that go through heavy occlusions, scale variations, and severe deformations. In [13] the author Mingwei Sheng et al., proposed a method for tracking the motion of the athlete–Distractor-aware SiamRPN (DaSiamRPN) network which is also used to track the result of the heavy rely..? Positions of the objects. And classification of the key frames is done using the Haar feature-based cascade. Finally, by the combination of HOG and a linear SVM the athlete of the sports video is detected. Athlete tracking is most important in recent studies to solve many more problems in sports and also to protect the future of sports.

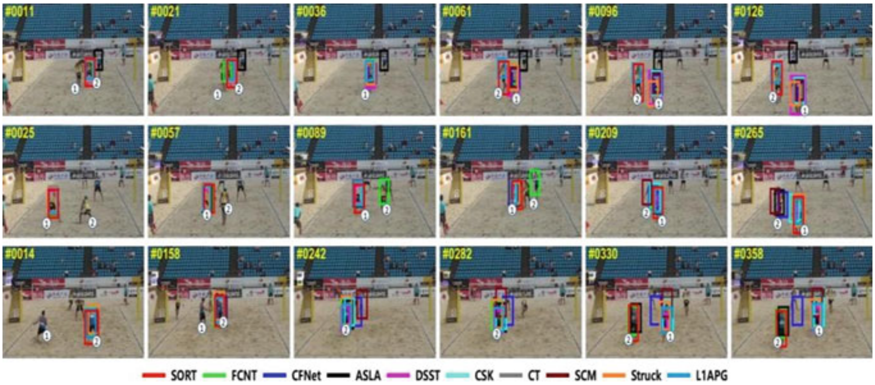


Fig. 2 BeaVoll dataset which goes through heavy occlusions, large scaling, and deformations [6]

In many sports like soccer, it would be difficult to identify players having the same clothes of the same colors for indexing and retrieving the applications. So, the proposed study describes the player identity in sports video with suitable algorithms. The most common problem in the existing methods is not considering player identity specifically. In [5] taking that into consideration the author proposed three parts of the framework that consists of the DeepPlayer model which extracts the features of the number and partial feature embedding. Secondly, 3D localization with the ID of the players uses an Individual Probability Occupancy Map (IPOM) model, and lastly, a proposed player ID links the nodes in the flow graph based upon the K-Shortest Path with an ID can be said to be as (KSP-ID) model. The architecture of the DeepPlayer model is shown in Fig. 3. The distinguished identity helps in improving the performance tracking. In [7] Long- teng Kong et al., proposed a novel approach to solve the hierarchical deep association of the same identity detection issues in sports videos. The encoded powerful deep features are used to employ the detection association. And, the new deep architecture of Siamese Tracklet Affinity Networks (STAN) presented for the tracklet association helps to model the long-term dependencies between the athletes. Finally, the minimum-cost network flow algorithm is used to solve the hierarchical association. Many coaches think that players' identity tracking would help to track both technical performance and health related performance.

2.2 *Sports Video Action Recognition*

This section describes the analysis of sports videos with various learning techniques that overcomes the issues like frequent occlusion, background clutter, out of view, motion blur and Inter-class similarity. According to the author [1] many methodologies predicting video sequences either omit or could not use the temporal information

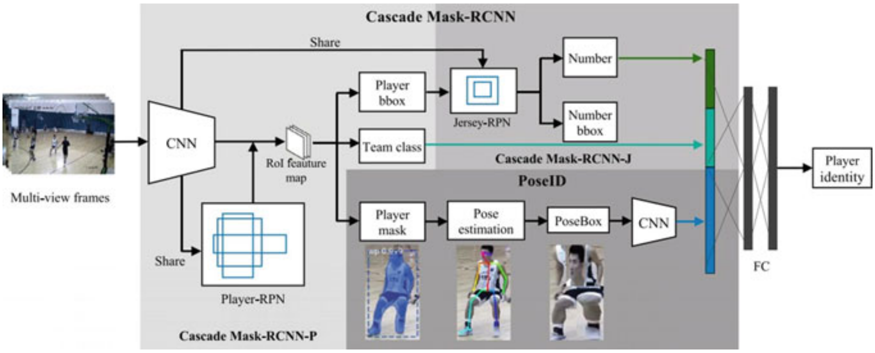


Fig. 3 The architecture of the DeepPlayer model [5]

for action recognition and also to reduce the high computational cost of human action recognition. To overcome this issue the author proposed a preprocessing phase with the combination of background subtraction, HOG, and Skeletal models that are used to analyze and illustrate the suitable frames. Later, the grouping of the CNN [23] and LSTM [21] networks was implemented for the feature selection process. At last human activities are labeled using the Softmax KNN classification method. The above methods are evaluated based on the UCF101 dataset [39]. In [2] the author proposed a work to overcome the satisfied accuracy problems in real-time online data stream processing. The author implements an efficient and optimized process from the surveillance environment of visual sensors based on CNN. Initially, the frame-level features are extracted using the pre-trained CNN model. Then the system is followed by the optimized deep autoencoder (DAE) method to study the temporal changes of actions. Finally, the SVM classifies the human action and the fine-tuning procedure is added to the testing phase, using accumulated data. It is believed that the planned method is appropriate for action recognition in surveillance data streams [52]. The Model Proposes a method for feature extraction using RCNN [3]. In this work, the author proposes a method for extracting the features and analysis of video content. Gaussian mixture model along with Kalman filter used to track the motion and other features are extracted by means of RNN [24] with the gated recurrent unit. This novel approach takes advantage of analyzing and extracting all features every time and in every frame of the video. Here [4], the several inclusive pre-trained models undergo hybrid approaches that can be achieved by Meta heuristic and genetic algorithms. This algorithm's major work is to merge the features obtained. Later, merging the features of the hybrid model is first compared with each model and then with all the scenarios. Then, the state of the art studies is compared to experiments with the performance optimization results. From the above study it is clear that CNN and LSTM play a major role in human action recognition from surveillance videos.

In [8], the author proposed the motion expression ability to improve, based on the spatiotemporal features that were first designed for the important key frame algorithm of motion videos.

Next, only the player is tracked and the interference fields in the video are omitted as referred to in Fig. 4. Finally, the behaviors of the players are classified. The classification of the video keyframe feature sequences is done with the help of CNN and RNN frameworks. Next, to overcome the challenges like low efficiency and high error rate in basketball action recognition, [9] the author proposed a work going on basketball shooting action that depends on feature extraction in addition to machine learning. This method elaborates the study of image feature extraction and Gaussian hidden variables that are used to recognize the basketball shooting gesture. This paper [10] uses the visual attention mechanism as well as two-stream attention based LSTM network. Furthermore, taking into consideration the correlation among two deep feature streams, a deep feature correlation layer is planned to regulate the deep learning system parameter. In [11] the main aim of the author is to build an automated solution to evaluate the handball player's motions. Here, the new dataset named RGB-D dataset is used. The player actions are filmed to learn who performed similar actions, and depth data and skeletons are sensed using the Kinect V2 sensors. The main actions of the players are examined using skeleton data simulation and they use dynamic time warping techniques to balance with the action among two players. This enables the coach to identify the player's performance and train the players accordingly.

In [12] taking action recognition in basketball as a challenge, the authors Zhiguo Pan and Chao Li proposed a novel model to recognize human action in basketball. The dataset of the basketball sports video is taken from YouTube and manually labeled. The feature from the motion regions is calculated by the Gaussian mixture model (GMM) followed by the gradient histogram being implemented to represent the shape descriptor. Finally, the motion and shape descriptor are combined and using the KNN algorithm the action recognition of basketball is determined [14]. In this model the author used a LSTM network to symbolize an individual person's action dynamic in sequence and another model aggregates the person-level details. Both the collective activity dataset and the new volleyball dataset are evaluated. In

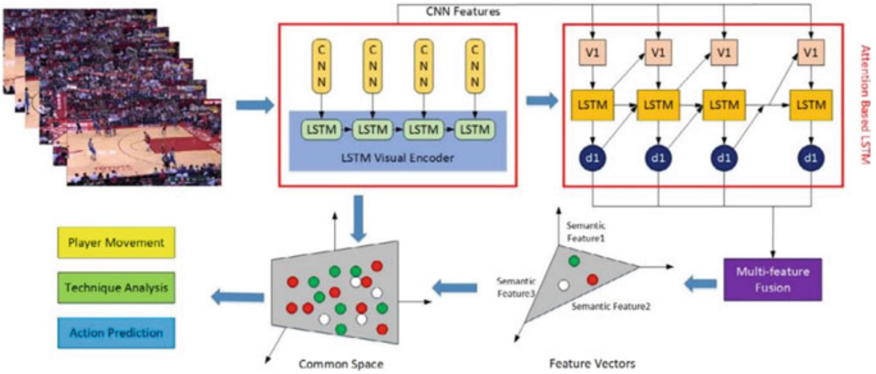


Fig. 4 The encoder-decoder based video motion recognition framework [8]

[15] the author uses his own hockey dataset to recognize hockey activity. Free hit, goal, penalty corner and long corner are the four main activities that are included in the hockey dataset. The LSTM [22] model is used to extract the spatial feature and deep learning model VGG-16 a pre-trained model is implemented for hockey activity recognition. In paper [16], Silvia Vinyes Mora and William J. Knottenbelt, proposed a method for action recognition in a tennis sports video deep learning approach. Here, the three layered long and short-term memory (LSTM) model is presented to work on the classified fine-grained tennis actions. There are not many studies based on badminton sports video analysis; so in [17] Nur Azmina Rahmad et al., decided to propose a method based on badminton sports. Here, the broadcast video of badminton is taken and Faster Region Neural Network is implemented to follow the positions of the players.

The author in [18] introduced a new dataset for a soccer game that consists of 222 broadcast videos. All the 222 videos cover the major annotations of soccer events, which are shoot annotation, event annotation and story annotation. Deep learning techniques such as Long-term short memory model and VGG model are implemented to extract the feature and detect the event of the soccer game. In [19] Gun Junjun, the dataset contains footnotes of detailed actions in a video. The data are classified using the adaptive multi-label classification methods. The main aim of the proposed work is to track the common interest and links with the group of individuals in the model. For this purpose, an FPGA network is used that identifies and extracts the features. The Ref. [20] proposed a work with the combination of convolutional neural network along with recurrent neural network works in the classification part. Then, the extracted features of CNN are combined with the RNN temporal information. Finally, the VGG- 16 model is applied with transfer learning to achieve performance accuracy classification. The above study explains various techniques used to recognize and track the in- formation in sports videos. Majorly in [8, 10, 14–16, 18, 20] references, CNN [22] and RNN [24] frameworks and LSTM [21, 25] methods are used for the recognition and tracking process.

2.3 Dataset

This section explains the datasets used in existing studies and also some of the popular datasets that are used in sports videos. A detailed list is shown in Table 1. These datasets differ in camera motion, number of actions, blur background, etc. and are used for the evaluation of the assorted algorithm.

APIDIS dataset [5, 27], and STU dataset [5] are based on basketball sport in Fig. 5. It is known as a publicly available dataset; here, the data are recorded using 7 cameras with 25 fps in 800*600 resolutions. It includes 1500 frames with 16 unlabeled periods of data additionally. STU dataset is a novel dataset composed from Shantou University. The videos are recorded with 8 cameras at 24 fps in 1280*720 resolutions. This dataset contains 11 periods where only 2 periods are evaluated in

Table 1 List of available datasets used for sports video analysis

References	Datasets	Year	Source	Sports
[5]	STU	2020	Stantou university	Basket ball
[6]	BEAVOLL	2019	General administration tradition of sports in china	Beach volley ball
[7]	VolleyTrack	2020	YouTube	Volley ball
[10]	UCF11	2020	Open Source	All
[10]	UCFSports	2020	Open Source	All
[10]	jHMDB	2020	Open Source	All
[14]	Volleyball	2016	YouTube	Volley ball
[15]	Hockey dataset	2020	YouTube	Hockey
[27]	APIDIS	2009	Open Source	Basket ball
[28]	NCAA basketball	2009	YouTube	Basket ball
[29]	H3DD	2019	Open Source	Hockey
[30]	Thetis	2013	Open Source	Tennis
[31]	HMDB	2011	Open Source	Tennis
[32]	Soccer-8 k	2019	YouTube	Soccer
[33]	Soccer dataset	2009	Surveillance	Soccer
[34]	Soccer dataset	2019	Surveillance	Soccer
[35]	SoccerNet	2018	Open Source	Soccer
[36]	RobocupSimData	2017	Open Source	RoboCupSoccer
[37]	SoccerNet-V2	2021	Open Source	Soccer
[38]	GolfDB	2019	Open Source	Golf

the proposed work. Both APIDIS and STU datasets are mainly used for the player's identity in [5].

BEAVOLL dataset [6] is a new benchmark dataset (in Fig. 6) and is based on beach volleyball games which are larger and very difficult datasets; it provides action labels for determining players' tracking and action recognition methods. It consists of 30 video clips that last from 80–120 per second. The videos are captured with 1440*1080 resolutions. To confirm action recognition, the dataset was created based on the 9 typical actions, like serve, dig, non-action, pass, spike, block, save, walk and run.

NCAA Basketball [28], VolleyTrack [7]: NCAA Basketball datasets are created based on the YouTube videos for team action recognition. The sequence consists of 1179 frames with 30fps in 640*480 resolutions. Since it is small and not enough to train the network, the APIDIS dataset [27] is combined with the NCAA basketball dataset for testing and training in the proposed model. VolleyTrack is a newly collected dataset that contains 18 video clips collected from YouTube. Each video lasts for about 8–12 s that include 5406 frames at 30fps with 1920*2080 resolutions. Both the datasets are shown in Fig. 7.

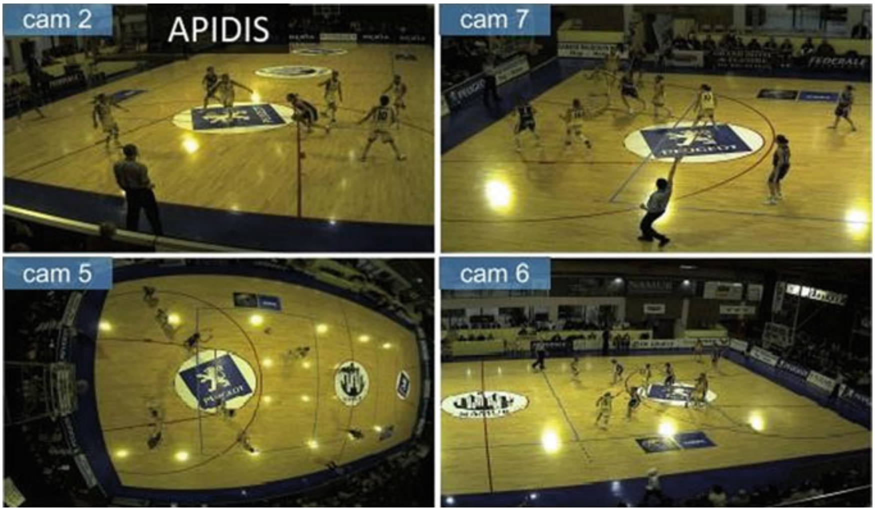


Fig. 5 Sample APIDIS dataset [7]

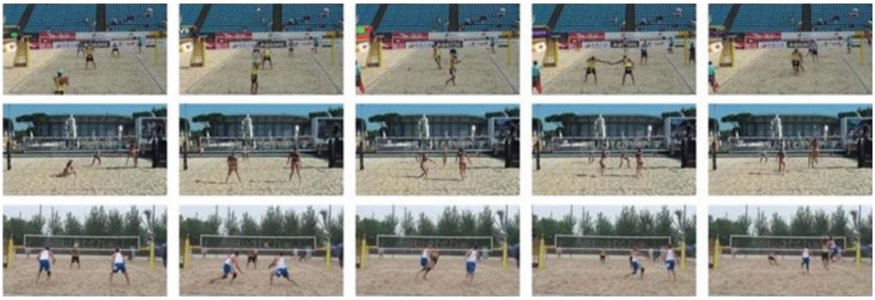


Fig. 6 Sample BEAVOLL dataset [6]

UCF11, UCFSports, jHMDB [10] UCF11 dataset contain 1600 videos with 11 different actions. Due to large variations of illumination, it is known as a challenging dataset [26]. UCFSports dataset is collected from broadcast television channels. It has 150 sequences with 720*480 resolutions are included in this dataset. jHMDB is a larger dataset that contains 923 videos with 21 various actions.

H3DD [29] dataset is an open source. The dataset contains handball activities in 3d video sequences. For every individual player RGB frame, depth frame and skeleton frame are noticed. According to the resolution size, the frames are categorized; the RGB data has the frame size of 1920*1980, the depth data has the frame size of 514*424 and a.txt file has skeleton data with 3D coordinates for 25 joints.

Hockey dataset [15] is a new dataset that was gathered from YouTube that contains the videos of Hockey World Cup 2018 and International Hockey Federation. The resolution of the video is 1280*720 size. The hockey dataset includes actions like

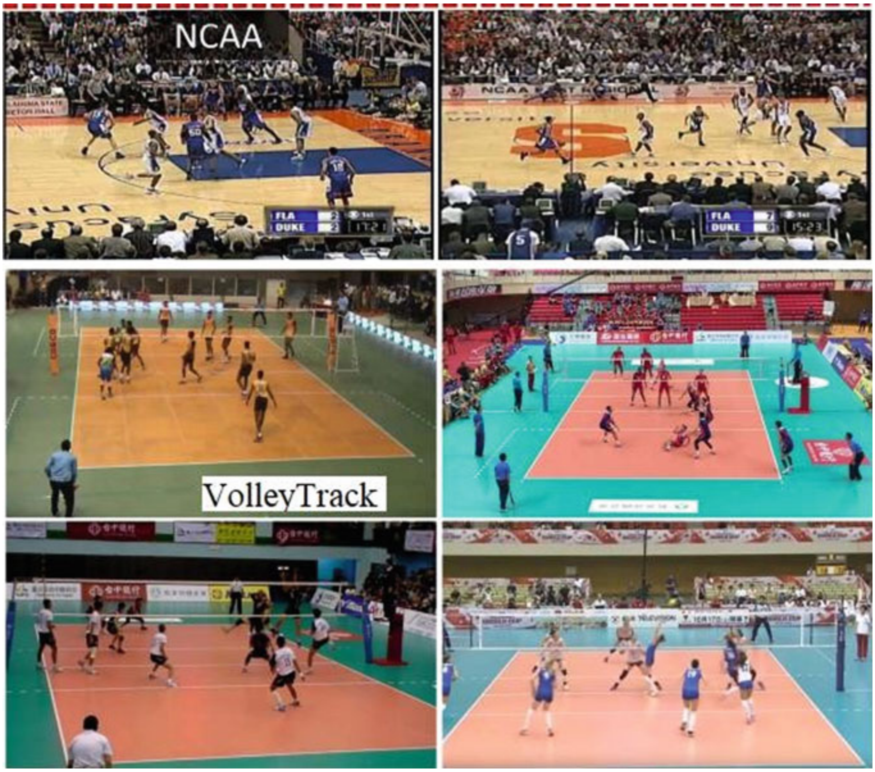


Fig. 7 Sample NCAA and VolleyTrack Datasets [7]

free hit, long corner and penalty corner of video frames, which are collected from 12 broadcasted hockey matches.

THETIS [16, 30], HMDB [16, 31]: THETIS dataset is created with 1980 tennis videos based on 12 actions. Each action is performed several times with 31 amateurs and 24 experienced players. This dataset is considered to be a longer dataset. HMDB dataset is used for action recognition. This dataset contains 6849 videos with various 51 actions that include facial actions along with body movements.

Soccer-8 k Dataset [32, 33] there is no standard dataset available for soccer games. Here, with the help of an event detector and based on the full HD videos of Laliga and a few Champions League the soccer -8 k dataset is created, and shown in Fig. 8. All the above datasets are used in the various applications of sports video analysis for both athlete tracking as well as action recognition; each dataset helps in a unique way to recognize actions in the sports video analysis.

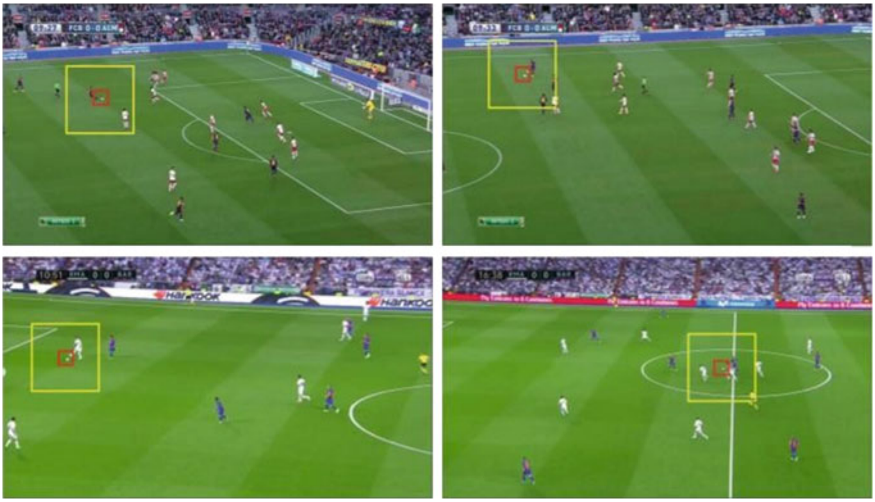


Fig. 8 Sample Soccer-8 k Dataset with correct event detection [32]

2.4 Advantages and Disadvantages of Action Recognition in Sports Video

From the above literature reviews the merits and the demerits of sports video action recognition are listed below,

Advantages

- Visualizing the function and understanding the representation is simple and easy for analysis
- The skeleton features from the data are more important.
- The model of the features that are used for training the data is openly known.
- Through the machine learning algorithms, the features are robotically learned.
- The deep learning algorithms are more adoptable for solving complex tasks.

Disadvantages

- Due to excessive dimensions, they may cause computational intensive.
- Human Detection in occlusion.
- Difficulty in updating the background information at regular intervals.
- Accurate identification of human motion is not recognized.
- Fixed-angle camera.
- Model needs to be trained for recognizing sub-types of action.
- Long-term dependencies between the frames.

3 Comparative Analysis

The below table comparatively summarizes about the deep learning algorithms used for action recognition in sports videos.

4 Result and Discussion

Analysis of sports performance is the investigation of real sports video performance. In [8] the movement recognition and movement prediction with its movement type results are presented in Table 2 as well as Table 3. The accuracy of a valid set and testing set of movement type pass/foul is lower for both recognition and prediction results. The proposed work recognition and prediction accuracy is compared with the existing method recognition and prediction accuracy of the datasets and shown in Fig. 9. The proposed work uses a larger dataset that gives high accuracy and identifies the players efficiently (Table 4).

In [10] the confusion matrix is used to make clear differences between the different datasets used in the proposed work. The above Fig. 10a and b shows the confusion matrix of the UCF11 dataset and UCF Sports datasets. The confusion matrix helps to identify the classes with the highest accuracy and lowest accuracy. In Fig. 10a class s-juggling has the lowest accuracy of 90.3% and class divine has the highest

Table 2 Comparative table of deep learning method

References	Year	Algorithm	Dataset	Accuracy (%)
Kong et al. [6]	2019	Long-term Recurrent Region-guided Convolutional	BeaVoll	73.2
Chen and Wang [8]	2020	CNN and RNN	NBA	76.5
Dai et al. [10]	2020	LSTM	UCF11 UCF Sports jHMDB	96.9 98.6 76.3
Rangasamy et al. [15]	2020	Pre-trained VGG-16 model	Hockey	98
Vinyes Mora and Knottenbelt [16]	2017	Three layered LSTM	THETIS	88.16

Table 3 Movement recognition result

Movement recognition result on the basketball player motion recognition dataset				
Movement type	Shoot (%)	Pass (%)	Dribble (%)	Foul (%)
Accuracy/valid set	84	75	90	80
Accuracy/testing set	82	70	85	74

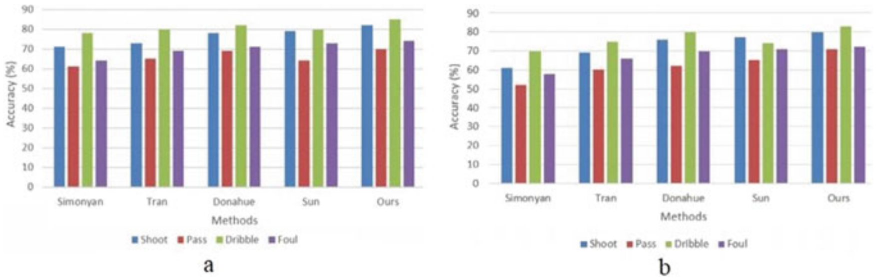


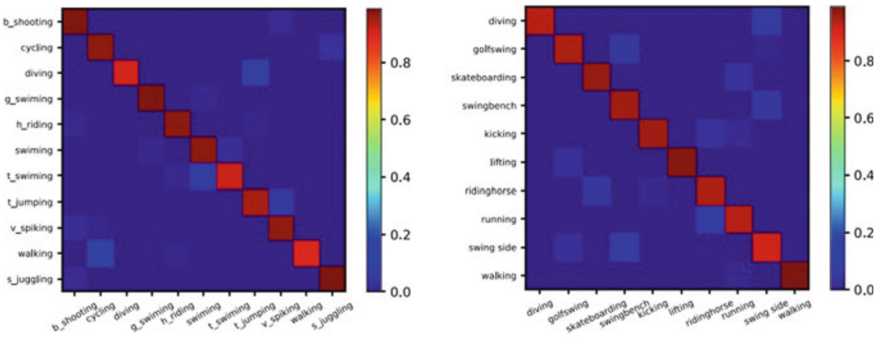
Fig. 9 **a** Represents the recognition accuracy for the test set and **b** Represents the prediction accuracy for the test set [8]

Table 4 Movement prediction result

Movement prediction result on the basketball player motion recognition dataset				
Movement type	Shoot (%)	Pass (%)	Dribble (%)	Foul (%)
Accuracy/valid set	83	73	87	77
Accuracy/testing set	80	71	83	72

ac- curacy of 99.8%. In Fig. 10b the accuracy is distributed uniformly and also all the accuracy classes have higher than 95% of accuracy, except the swing side class action. Figure 11 shows the confusion matrix of the proposed work of the jHMDB dataset. Here, some actions have higher accuracy and some have lower accuracy comparatively.

In [11] to calculate the potential level of players the set of performances had been made between the experts and beginners. To analyze their performance the author used the skeleton method of 3D motion. In [13] the proposed work uses testing datasets that include two video sequences where the video sequence I with 510 frames is the men's 100 m race video at Rio 2016 and the video sequence II with 380 frames is the London 2012 Olympic Games. Both the video sequences are



(a) Confusion matrix of our proposal for UCF11 dataset

(b) Confusion matrix of our proposal for UCF Sports dataset

Fig. 10 Confusion matrix of UCF11 and UCF Sports datasets [10]

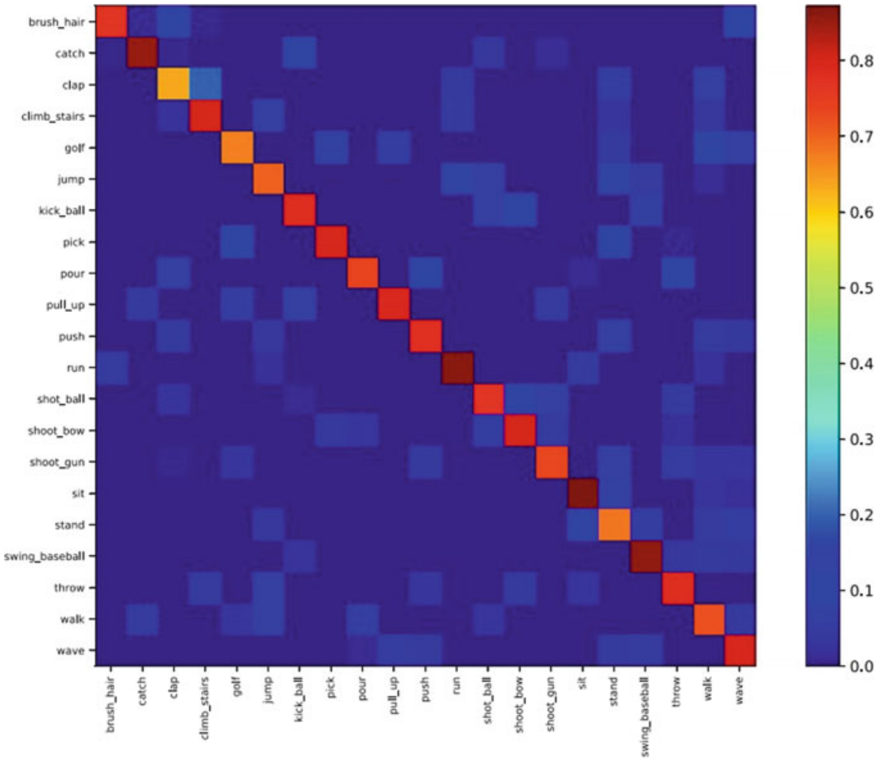


Fig. 11 Confusion matrix of jHMDB dataset [10]

combined together as 890 frames and proceeds for implementation. From the above section, it is decided that the pre-trained model gives better accuracy compared with other algorithms.

5 Inference Made

- An additional reference frame is included for occluded regions in a frame for accurate human detection. Occlusion can be reduced by the selection of camera positions.
- Background Subtraction of the sports video are compared and analyzed in dynamic scenes.
- Creating bounding boxes and identifying everything within each box is to improve the accuracy and precision of object detection and tracking in a video.
- Models need to be trained for recognizing subtypes of action.
- Multi-camera tracking systems were essential for tracking all the players.
- Lack of Short tracklets (frames).

6 Suggestion Proposed

The association of data analytics and sensors in sports video analysis provides needful information for players. Sensors are used to generate data that is essential in the field of the player's health, detects the player's posture, and fan engagement, by using sensor data transforming it with the real streaming analysis and then joining the data with artificial intelligence. This creates an emerging application developed in various fields. In future, predictive intelligence will help the sports team to improve the game strategies by coaching the players of the game based on athlete tracking. It is also used for extracting and analyzing players and team statistics. Using novel deep learning algorithms the system could help the teams and coaches improve the accuracy. And by training, the system will track the action of the players and their mistakes. Also based on the tracking details, the players are coached and monitored for better performance in future. The emerging techniques of deep learning can also be focused on predicting or analyzing the game results based on the audience's facial reaction. Later, based on the novel computer vision application the shots of the ball are predicted, as to whether it is in/out of boundary lines. In many existing works, the pre-recorded data streaming video is used; instead, the live video streaming data can be used for sports video action recognition. The training duration of the system could be reduced for performing with better accuracy.

7 Conclusion

This paper focuses on the athlete action recognition of sports videos; so it highlights some of the common issues in action recognition of sports, like athlete tracking and player identity detection. And also reviews its commercial applications that include video abstraction, performance analysis and augmented reality. A detailed survey of the emerging system of sports video action recognition and athlete tracking is highlighted. Along with it the availability of benchmark datasets is studied, that is very significant. Here, some datasets are globally available and their results are evaluated. But some evaluations are based on their own dataset which is not available globally.

And also the analysis of the sports video action recognition is experimented followed by the inference made. From the above studies it is concluded that the pre-trained model gives better accuracy in action recognition of the sports videos. An apparent future direction for sports video athlete action recognition is therefore imposed on different algorithms.

References

1. Serpush F, Rezaei M (2021) Complex human action recognition using a hierarchical feature reduction and deep learning-based method. SN Computer Science 2(2):1–15
2. Ullah A, Muhammad K, Haq IU, Baik SW (2019) Action recognitionusing optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments. Fut Gener Comput Syst 96:386–397.
3. Jaouedi N, Boujnah N, Bouhlel MS (2020)A new hybrid deep learning model for human action recognition. J King Saud Univ-Comput Infor m Scie 32(4):447–453
4. Yilmaz AA, Guzel MS, Bostanci E, Askerzade I (2020) Anovel action recognition framework based on deep-learning and genetic algorithms. IEEE Access 8:100631–100644
5. Zhang R, Wu L, Yang Y, Wu W, Chen Y, Xu M (2020) Multi-cameramulti-player tracking with deep player identification in sports video. Pat tern Recognit 102:107260
6. Kong L, Huang D, Qin J, Wang Y (2019) A joint framework for athlete tracking and action recognition in sports videos. IEEE Trans Circuits Syst Video Technol 30(2):532–548
7. Kong L, Huang Di, Wang Y (2020) Long-term action dependence-based hierarchical deep association for multi-athlete tracking in sports videos. IEEE Trans Image Process 29:7957–7969
8. Chen L, Wang W (2020) Analysis of technical features in basketball video based on deep learning algorithm. Signal Process: Image Commun 83:115786
9. Ji R (2020) Research on basketball shooting action based on image feature extraction and machine learning. IEEE Access 8:138743–138751
10. Dai C, Liu X, Lai J (2020) Human action recognition using two-stream at- tention based LSTM networks. Appl Soft Comput 86:105820
11. Elaoud A, Barhoumi W, Zagrouba E, Agrebi B (2020) Skeleton-based comparison of throwing motion for handball players. J Ambient Intell Humaniz Comput 11(1):419–431
12. Pan Z, Li C (2020) Robust basketball sports recognition by leveraging motion blockestimation. Signal Process: Image Commun 83:115784
13. Sheng M, Wang W, Qin H, Wan L, Li J, Wan W (2020) A novel changing athlete body real-time visual tracking algorithm based on distractor-aware SiamRPN and HOG-SVM. Electronics 9(2):378
14. Ibrahim MS, Muralidharan S, Deng Z, Vahdat A, Mori G (2016) A hierarchical deeptemporal model for group activity recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1971–1980
15. Rangasamy K, As'ari MA, Rahmad NA, Ghazali NF (2020) Hockey activity recognition using pre-trained deep learning model. ICT Express 6(3):170–174
16. Vinyes Mora S, Knottenbelt WJ (2017)Deep learning for domain-specific action recognition in tennis. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 114–122
17. Rahmad NA, Sufri NA, Muzamil NH, As'ari MA (2019) Badminton player detection using faster region convolutional neural network. IndonesianJ Electr Eng Comput Sci 14(3):1330-1335
18. Yu J, Lei A, Hu Y (2019) Soccer video event detection based on deep learning. In: Internationalconference on multimedia modeling. Springer, Cham, pp 377–389
19. Junjun G (2021) Basketball action recognition based on FPGA and particle image. Microprocess Microsyst 80:103334
20. Russo MA, Kurnianggoro L, Jo KH (2019) Classification of sports videos with combination of deep learning models and transfer learning. In: International conference on electrical,computer and communication engineering (ECCE). IEEE, pp 1–5
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
22. Baccouche M, Mamalet F, Wolf C, Garcia C, Baskurt A (2010) Action classification in soccer videos with long short-term memory recurrent neural networks. In: International con ference on artificial neural networks. Springer, Berlin, pp 154–159
23. Yamashita R, Nishio M, Do RK, Togashi K (2018) Convolutional neural networks: anoverview and application in radiology. Insights Imaging 9(4):611–629

24. Tarwani KM, Edem S (2017) Survey on recurrent neural network in natural language processing. Int J Eng Trends Technol (IJETT) 48(6)
25. Lindemann B, Müller T, Vietz H, Jazdi N, Weyrich M (2021) A survey on long short-term memory networks for time series prediction. Procedia CIRP 99:650–655
26. Liu J, Luo J, Shah M (2009) Recognizingrealistic actions from videos "in the wild". In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 1996–2003
27. De Vleeschouwer C, Delannay D (2009) Basketball dataset from the European project APIDIS
28. Dataset, https://www.kaggle.com/ncaa/ncaa-basketball
29. Dataset, https://github.com/Elaoud/H3DD-dataset
30. Gourgari S, Goudelis G, Karpouzis K, Kollias S (2013) Thetis: three dimensional tennisshots a human action dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 676–681
31. Kuehne H, Jhuang H, Garrote E, Poggio T, Serre T (2011) HMDB: a largevideo database for human motion recognition. In: 2011 international conference on computer vision. IEEE, pp 2556–2563
32. Ganesh Y, Sri Teja A, Munnangi SK, Murthy GR (2019) A novel framework for finegrained action recognition in soccer. In: International work-conference on artificial neural networks. Springer, Cham, pp 137–150
33. D'Orazio T, Leo M, Mosca N, Spagnolo P, Mazzeo PL (2009) A semi-automatic system forground truth generation of soccer video sequences. In: 6th IEEE international conference on advanced video and signal surveillance. Genoa, Italy September, 2–4
34. Kukleva A, Khan MA, Farazi H, Behnke S (2019) Utilizingtemporal information in deep convolutional network for efficient soccer ball detection and tracking. In: Robot World Cup. Springer, Cham, pp 112–125
35. Giancola S, Amine M, Dghaily T, Ghanem B (2018) Soccernet: a scalable dataset for action spotting in soccer videos. In: Proceedings of the IEEE conference on computer visionand pattern recognition workshops, pp 1711–1721
36. Dataset: https://paperswithcode.com/dataset/robocupsimdata
37. Deliege A, Cioppa A, Giancola S, Seikavandi MJ, Dueholm JV, Nasrollahi K, Ghanem B, Moeslund TB (2021) Soccernet-v2: a dataset and benchmarks for holistic understanding of broadcast soccer videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4508–4519
38. Dataset,https://github.com/wmcnally/GolfDB
39. Soomro K, Zamir AR, Shah M (2012) UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402
40. Rui Y, Gupta A, Acero A (2000) Automatically extracting highlights for TV baseball programs. In: Proceedings of the eighth ACM international conference on Multimedia, pp 105-–115
41. Kawashiima T, Yoshino K, Aok Y (1994) Qualitative image analysis of group behavior. In: Proceedings computer vision and pattern recognition
42. Taki T, Hasegawa JI, Fukumura T (1996) Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In: Proceedings of 3rd IEEE international conference on image processing, vol 3. IEEE, pp 815–818
43. Yu X, Leong HW, Xu C, Tian Q (2004) A robust and accumulator-free ellipse Houghtransform. In: Proceedings of the 12th annual ACM international conference on multimedia, pp 256–259
44. Yu X, Yan X, Hay TS, Leong HW (2004)3D reconstruction and enrichment of broadcast soccer video. In: Proceedings of the 12th annual ACM international conference on multi- media, pp 260–263
45. Bebie T, Bieri H (2000)A video-based 3D-reconstruction of soccer games. Eurographics vol 19(3)
46. Pingali G, Opalach A, Jean Y, Carlbom I (2001) Visualization of sports using motion trajectories: providing insights into performance, style, and strategy. In: Proceedings visualization,2001. VIS'01. IEEE pp. 75–544
47. Rui Y, Gupta A, Acero A (2000) Automatically extracting highlights for TV baseball programs. In: Proceedings of the eighth ACM international conference on multimedia, pp 105-115

48. Ahammed Z (2018) Basketball player identification by jersey and number recognition (Doctoral dissertation, Brac University)
49. Lu CW, Lin CY, Hsu CY, Weng MF, Kang LW, Liao HY (2013) Identificationand tracking of players in sport videos. In: Proceedings of the fifth international conference on internet multimedia computing and service, pp 113–116
50. Niebles JC, Chen CW, Fei-Fei L (2010) Modeling temporal structure of decomposablemotion segments for activity classification. In: European conference on computer vision. Springer, Berlin, pp 392–405
51. Kausalya K, Chitrakala S (2012) Idle object detection in video for banking ATM applications. Res J Appl Sci Eng Technol 4(24):5350–5356
52. Sandhiya B, Priyatharshini R, Ramya B, Monish S, Raja GR (2021)Reconstruction, identification and classification of brain tumor using gan and faster regional-CNN. In 2021 3rd international conference on signal processing and communication (ICPSC). IEEE.