TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 4, August 2020, pp. 1926~1933
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i4.14730

Deep learning in sport video analysis: a review

Keerthana Rangasamy[1], Muhammad Amir As'ari[2], Nur Azmina Rahmad[3], Nurul Fathiah Ghazali[4], Saharudin Ismail[5]
[1,2,3,4,5]School of Biomedical Engineering and Health Sciences, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
[2]Sport Innovation and Technology Center, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, Malaysia

Article Info

Article history:

Received Dec 2, 2019
Revised Mar 2, 2020
Accepted Mar 13, 2020

Keywords:

Deep learning
Human activity recognition
Sport video analysis

ABSTRACT

Sport is a competitive field, where it is an element of measurement for a countries development. Due to this reason, sport analysis has become one of the major contribution in analysing and improving the performance level of an athlete. Video-based modality has become a crucial tool used in sport analysis by coaches and performance analysis. There were wide variety of techniques used in sport video analysis. The main purpose of this review paper is to compare and update review between traditional handcrafted approach and deep learning approach in sport video analysis based on human activity recognition, overview of recent study in video based human activity recognition in sport analysis and finally concluded with future potential direction in sport video analysis.

Corresponding Author:

Muhammad Amir As'ari,
Department of Biomedical Engineering,
School of Biomedical Engineering and Health Sciences,
Faculty of Engineering,
Universiti Teknologi Malaysia,
Johor Bahru, Malaysia.
Email: amir.asari@biomedical.utm.my

1. INTRODUCTION

Sport is an activity that involves the active movement of an athlete either in the form of a team or individual to complete with the opposite team [1]. It is one of the elements of measurement for a countries development as it brings a good reputation to a country if its team win in the game [2]. In order to achieve high performance in sport by the athletes, coaches and sports professionals play a vital role to evaluate and train their athletes [3]. Among various methods, sport video analysis is one of the methods to evaluate the performance level of athletes and to enhance training techniques [4]. A process to assess the performance level of an athlete is known as performance analysis [5]. Performance analysis can be divide into two which are technical analysis and tactical or notational analysis [6]. Through technical analysis, we would get an answer to the question: how does the game is performed by the athletes. On the other hand, through tactical or notational analysis the question of what activity is performed would be answered [7].

Over the past few years, in the field of computer vision, different approaches have been implemented to analyse sport videos. During the early stage traditional handcrafted approaches were proposed by many researchers for sport video analysis [8]. After the development of technology, deep leaning approach has become the current topic of interest in sport video analysis due to its successful

performance in other computer vision areas such as human-computer interface, handwriting recognition and speech recognition [9]. Generally, sport video analysis using such approach can be divide into several directions such as tracking players as well as ball, trajectory detection and action recognition. This paper aims to review current trends in deep leaning approach in human activity recognition-based sport video analysis. The remainder of this paper is organized as section 2 illustrates traditional handcrafted architecture. Next section 3 presents deep learning architecture. Section 4 discusses the drawback of state of the art works and Section 5 present conclusion and future works to address.

2. TRADITIONAL HANDCRAFTED ARCHITECTURE

In this section, overview of handcrafted architecture and its application in sport video analysis was studied based on prior research.

2.1. Overview of handcrafted architecture

Before the advancement of technology and the emergence of deep learning architecture, sport video analysis is mainly performed using handcrafted features [9]. Traditional handcrafted features are human design features for specific problems [10]. It consists of feature descriptors and feature extractors that used to manually extract the important information from the sport video for further analysis. However, these traditional handcrafted features could only execute low-level features extraction [11]. Further in this section would like to present the existing traditional handcrafted architecture that was used widely in computer vision to recognize activity in video.

During the prehistoric ages of sport video analysis, action recognition was performed using manually human design features to extract and represent features for action recognition by which this technique is called a handcrafted approach. At first, low-level action features such as histogram of optical flows (HOF), histogram of gradients (HOG), sparse spatial-temporal interest points (STIP) features were designed to extract features in videos for analysis [9]. Then, those sparse features were embellished to dense spatio-temporal features [12]. After some time, improved dense trajectories (iDT) were innovated which comprises optical flow and speeded up robust features (SURF) [13]. To further enhance iDT multi-layer staked fisher vector (FV) was developed [14]. Multi-skip Feature staking was also used for feature extraction of action recognition in videos which improve the performance of action recognition [15]. Although, these handcrafted features were evolved with time and also improved in performance but it could only perform in a specific problem. It is not flexible to be used in other datasets. Table 1 shows some of the existing popular handcrafted features used in action recognition for video analysis such as sport video analysis.

Table 1. List of popular handcrafted features used for action recognition in videos

Handcrafted features	Reference
Histogram of Optical Flow (HOF) and Histogram of Gradients (HOG)	[16, 17]
Sparse Spatio-temporal interest points	[18]
Improved Dense Trajectories (iDT)	[13]
Fish Vector (FV)	[14]
Multi-layer Feature Stacking (MLFS)	[15]

2.2. Sport video analysis using handcrafted features

During the early stages of video analysis in computer vision such as in sports, handcrafted features were used as mentioned earlier. For instance, Abdulmunem et al. [19] were proposed a method for action recognition based on salient object spotting using local and global descriptors in KTH and UCF-Sports datasets. In this method, they were using handcrafted 3D-SIFT-HOOF (SGSH) features which were pass through SVM after encoding with bag of visual words approach. Besides, Carnonneau et al. [20] presented a novel technique to identify play break events in hockey videos. They used bag-of-words event detectors to identify the key events like line-change and play-break STIP were used in final decision making by creating context descriptor. Figure 1 the shows schematic block diagram of the event detector used in [20].

Zhu et al. [7] introduced a novel method to detect an event in a soccer games to extract tactical information. In this research work, they employ multi-object trajectory and field locations of the event shots to recognize the semantic event in the soccer games. Moreover, Lee et al. [21] implement pattern matching techniques to automatically recognize events in soccer games using a multi-object tracking unit and motion recognizer. Their proposed block diagram is shown in Figure 2.

Lien et al. [22] studied scene-based event detection for baseball videos by implementing various handcrafted features such as image-based features, object-based features and global motion to capture

semantic information from the baseball sport video. Then those, extracted features were fed into hidden Marvon model (HMM) to classify the detected events. Chen et al. [23] developed a novel handcrafted model based on player trajectories reconstruction method which includes court detection, player detection, player tracking and homography transformation in broadcasted basketball videos.

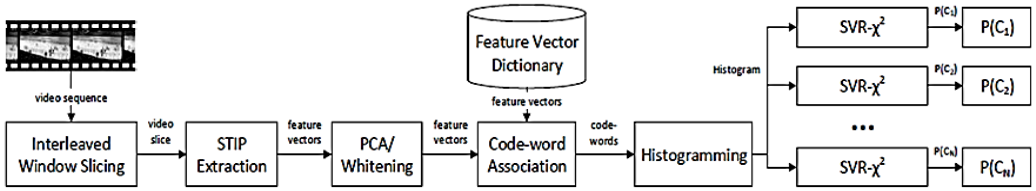

Figure 1. Schematic block diagram of event detectors in [20]

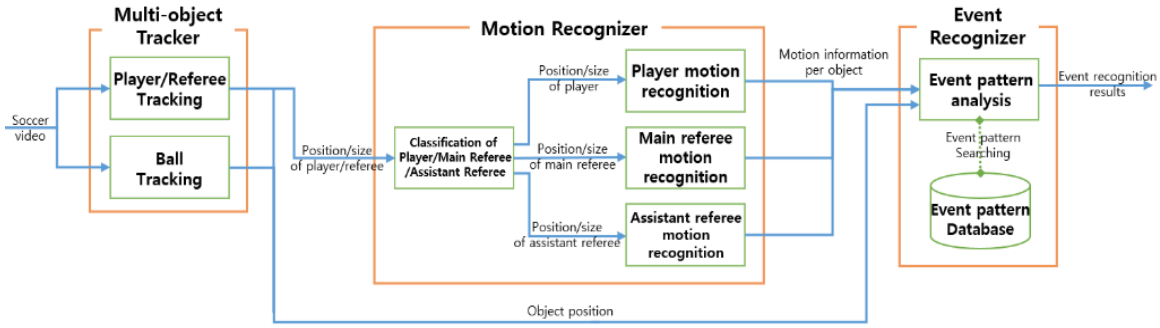

Figure 2. Block diagram of soccer event detection system [21]

Lu et al. [24] constructed a system to automatically track various hockey players as well as recognize their activity concurrently in broadcasted hockey video. In this study, HOG descriptors were utilised to detect the players and to model the appearance of the players a probabilistic framework was designed with a mixture of local subspaces. Additionally, boosted particle filter (BPF) was employed in this tracking system. Mukherjee et al. [25] proposed a novel descriptor based on improved dense trajectories in human action and event recognition. Their design also used Fisher Vector (FV) and the proposed novel approach was trained with binary support vector machine (SVM). The dataset used for this research was UCF sports, CMU Mocap and Hollywood2 datasets for event and action recognition.

3. DEEP LEARNING

In this section, overview of deep learning archicture and its application in sport video analysis was reviewed based on previous studies.

3.1. Overview of deep learning architecture

Deep learning architecture works automatically by directly classifying raw input images or video frames through multiple processing layers so as to learn and represent data [26]. Unlike traditional handcrafted architecture, it does not requires any feature descriptors or feature extractors. For instance, deep learning architecture uses local perception, down pooling, weight sharing, a multi-convolution kernel, etc. to automatically learn local features from just a segment of an image rather than whole image [27]**.** Deep learning techniques able to classify high-level or complex action recognition which attracts huge research of interest [28]. Examples of widely used deep learning models is convolutional neural network (CNN), recurrent neural network (RNN) and long short-term memory (LSTM).

Excellent performance with overwhelming accuracy of deep learning in visual task inspired the exploitation of deep learning in video analysis. Initially, CNN works independently to extract information from the still images [29]. However, 2D-CNN fails to extract temporal information in video sequences. In order to overcome this issue, 3D-CNN is then constructed to extract both spatial and temporal information of video frames [30]. Following this RNN was used in action recognition. RNN based method effectively

capture temporal information whereby its current prediction is based on both current observations as well as past information [31]. But in spite of that, RNN architecture only has short term memory, which could not apply in the real-world scenario. To alleviate this problem, LSTM model was proposed. This model able to extract temporal information from sequential video data. LSTM model has a memory unit that decides when to remember and forget hidden states [32]. Due to its superiority, the LSTM model broadly used in computer vision applications such as action recognition. Table 2 show a comparison between deep learning models.

Table 2. Comparison between deep learning model

Model	Advantage	Drawback
2D-CNN	▪ Automatically capture the spatial information in the image patches	▪ Could not able to capture the temporal ▪ information in video data
3D-CNN	▪ Automatically capture both spatial and temporal information	▪ Expensive model due to 3D
RNN	▪ Automatically capture the temporal information in sequential data	▪ Has short memory ability, could not apply in real situation ▪ Gradient explosion
LSTM	▪ Automatically capture the temporal information in sequential data	▪ NIL

3.2. Deep learning architecture in sport video analysis

Despite the astonishing performance of deep learning approach in various computer vision application such as voice recognition, text recognition it also achieves outstanding results in sport video analysis in recent years. Although it is still in the early stage of application in sport video analysis and only very few research has been done, yet so far its performance is more accurate as compared to traditional approaches and it's getting more attention presently [33]. Tora et al. [34] proposed a novel deep learning approach in classifying puck possession events in ice hockey. They used pre-trained CNN to first extract the features then use LSTM for classification of the five types of events which are dump in, dump out, loose puck recovery, pass and shot. Sozykin et al. [35] presented a 3D CNN based action recognition system for multi-class imbalanced in ice hockey. They first extract features from both single image and a slice of frames using CNN. Then, to solve the multi-class imbalance, they introduce two different strategies that can compare its performance separately which are the ensemble of *k* independent single-label learning networks and a single multi-label *k*-output network. It is found that as compared to the ensemble model, single multi-label *k* output model achieves high performance. Longteng et al. [36] designed a joint framework comprising of both athlete tracking and action recognition in sport videos using scaling and occlusion robust tracker with spatial pyramid pooling network (SPP-net) [37] and LSTM to extract motion, spatial and temporal features.

Jiang et al. [38] introduce deep learning based automatic soccer video event detection by using CNN and RNN. This paper focused on 4 types of event which are goal, goal attempt, corner and card. CNN model was used to extract image features. And RNN structure was used to capture temporal relation. However, in this paper three different types of RNN structures (traditional RNN model, LSTM model and gated recurrent unit (GRU) model) were used to determine the optimum model. LSTM is the best-performed model among those three RNN structures. Hong et al. [39] delivered deep transfer learning based end-to-end soccer video scene and event classification. Scene classification includes long view and close-up view while event classification includes corner, free-kick, goal and penalty. Their classification model used CNN models as well as transfer learning method. Whereas, Yu et al. [40] designed a deep learning-based soccer event detection that includes event detection as well as story generation refer to Figure 3 which begins with event clips and end with replay event. In this design CNN and LSTM were used.

Besides, in tennis sport, Mora et al. [41] have proposed a domain-specific deep learning action recognition method by utilising pre-trained CNN with three-layered LSTM model. This paper was aim to recognised fine-grained action in tennis sports. The deep LSTM network used in this research is able to learn high-level structures and provides high accuracy. Apart from that, Ibrahim et al. [42] innovated a deep model to extract dynamic temporal information in volleyball using LSTM models. The game state was deduced by capturing players' state through the two-stage LSTM model. Ramanathan et al. [43] established a deep leaning method to detect and categorised basketball events using RNN. This work used an attention model to detect key players from multi-person videos first. Then used CNN and LSTM to extract feature and detect an event in basketball.

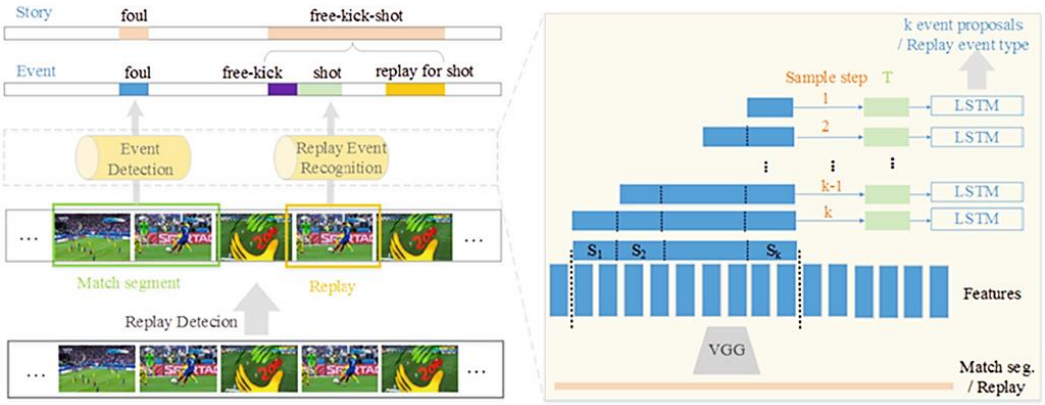

Figure 3. Proposed soccer story detection framework [40]

4. DISCUSSION

For the past few decades, there was quite a lot of research focusing on sport video analysis based on handcrafted architectures. However, this handcrafted architecture which consists of feature descriptors and extractors are problem-dependent. In other words, those handcrafted features can only apply to specific problems. It could not be used for other datasets or problems which is irrelevant to apply in a real scenario. Apart from that, the traditional handcrafted architecture used for sport video analysis could only extract low-level features, it is very challenging in capturing for high-level semantic information [44]. Thus, the advancement in technology leads to the emergence of deep learning-based sport video analysis. Since deep learning-based sport video analysis is still a new and growing research field, there were only a few studies found. Most of the studies are focusing on soccer games, tennis, baseball and basketball [45]. Only limited resources are found on the other sports such as hockey and badminton. However, the main drawback of using deep learning models are its data-hungry meaning to learn features automatically from raw inputs data it is depends on thoudsands of input data [46]. Besides, to elevate the performance level of the deep learning model it needs high-performance GPUs [47]. But, the recent advancement in technology and the growth in big data have overcome these issues.

Previously, CNN model which is one of the deep learning approach has shown tremendous success in image recognition, speech recognition audio recognition, etc [48]. However, in the analysis of video input data researchers face many challenges because video sequences dynamically evolve with time. It is difficult to extract temporal information. With the continuous study in video analysis using deep learning, cause the establishment of a sequential models such as RNN and LSTM. These models are able to extract temporal information in video input data. There were some researches work on the combination of both CNN and LSTM model to extract spatio-temporal information. But only a few researches were found in extraction high-level semantic information in sport video analysis. Despite astonishing performance of deep learning based archicture, the advancement achieves in image classification have not been reached in certain field like video classification or sport video analysis [49]. It is still an open issue in deep learning-based research in which many researchers try to solve and it is an ongoing research work [50].

5. CONCLUSION

This paper contributes a comprehensive survey on sport video analysis by comparing both handcrafted and deep learning approach. In summary, deep learning approach has overcome the limitations encountered by traditional methods in activity recognition of sport video analysis. However, only a few research has focused on sport video analysis. So, in future studies, the researchers can focus on extraction high-level semantic information in sport analysis which will be used by coaches and sports professionals in evaluating players' tactical performance in the game. Moreover, future research should also concentrate more on sport videos that are apart from soccer games, tennis, baseball and basketball as almost 80% of prior research had been focused on those sports.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Universiti Teknologi Malaysia (UTM) for endow this research and the Minister of Higher Education (MOHE), Malaysia for supporting this research work under Zamalah Scholarship and Research Grant No. Q.J130000.2545.19H88.

REFERENCES

[1] Lexico, "sport | Definition of sport in English by Oxford Dictionaries," [Online]. Available: https://en.oxforddictionaries.com/definition/sport. Accessed: 31 January 2019.

[2] D. Gu, "Analysis of tactical information collection in sports competition based on the intelligent prompt automatic completion algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 3, pp. 2927-2936, 2018.

[3] N. Homayounfar, S. Fidler, and R. Urtasun, "Sports field localization via deep structured models," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2017-January, pp. 4012–4020, 2017.

[4] M. Stein et al., "Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 13–22, 2018.

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

[6] N. A. Rahmad, M. A. As'ari, N. F. Ghazali, N. Shahar, and N. A. J. Sufri, "A survey of video based action recognition in sports," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, no. 3, pp. 987–993, 2018.

[7] G. Zhu et al., "Event tactic analysis based on broadcast sports video," *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 49–67, 2009.

[8] A. Kar, N. Rai, K. Sikka, and G. Sharma, "AdaScan: Adaptive scan pooling in deep convolutional neural networks for human action recognition in videos," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3376–3385, 2017.

[9] B. Meng, X. Liu, and X. Wang, "Human body action recognition based on quaternion spatial-temporal convolutional neural network," *Yi Qi Yi Biao Xue Bao/Chinese J. Sci. Instrum.*, vol. 38, no. 11, pp. 2643–2650, 2017.

[10] H. Yang et al., "Asymmetric 3D Convolutional Neural Networks for action recognition," *Pattern Recognit.*, vol. 85, pp. 1–12, 2019.

[11] Z. Huiqun, W. Hui, and W. Xiaoling, "Application research of video annotation in sports video analysis," *2011 International Conference on Future Computer Science and Education*, pp. 62–66, 2011.

[12] S. Dollár, P., Rabaud, V., Cottrell, G. and Belongie, "Going deeper into action recognition : A survey," *Image and Vision Computing*, vol. 60, pp. 4–21, 2017.

[13] H. Chen, J. Chen, R. Hu, C. Chen, and Z. Wang, "Action recognition with temporal scale-invariant deep learning framework," *China Communications*, vol. 14, no. 2, pp. 163–172, 2017.

[14] X. Peng, C. Zou, Y. Qiao, and Q. Peng, "Action Recognition with Stacked Fisher Vectors," *European Conference on Computer Vision - ECCV 2014: Computer Vision – ECCV 2014*, pp. 581–595, 2014.

[15] Zhenzhong Lan, Ming Lin, Xuanchong Li, A. G. Hauptmann, and B. Raj, "Beyond Gaussian Pyramid: Multi-skip Feature Stacking for action recognition," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 204–212, 2015.

[16] N. Dalal, B. Triggs, and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance," *European Conference on Computer Vision (ECCV '06)*, pp. 428–441, 2006.

[17] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[18] P. Dollár, V. Rabaud, G. Cottrell, S. Belongie,"Behavior recognition via sparse spatio-temporal features," *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2005.

[19] A. Abdulmunem, Y. K. Lai, and X. Sun, "Saliency guided local and global descriptors for effective action recognition," *Computational Visual Media*, vol. 2, pp. 97–106, 2016.

[20] M. A. Carbonneau, A. J. Raymond, E. Granger, and G. Gagnon, "Real-time visual play-break detection in sport events using a context descriptor," *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2808–2811, 2015.

[21] J. Lee, D. Nam, S. Moon, J. Lee, and W. Yoo, "Soccer Event Recognition Technique based on Pattern Matching," *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 643–646, 2017.

[22] C. C. Lien, C. L. Chiang, and C. H. Lee, "Scene-based event detection for baseball videos," *Journal of Visual Communication and Image Representation*, vol. 18, no. 1, pp. 1–14, 2007.

[23] L. H. Chen, C. W. Su, and H. A. Hsiao, "Player trajectory reconstruction for tactical analysis," *Multimedia Tools and Applications*, vol. 77, pp. 30475–30486, 2018.

[24] W. L. Lu, K. Okuma, and J. J. Little, "Tracking and recognizing actions of multiple hockey players using the boosted particle filter," *Image and Vision Computing*, vol. 27, no. 1, pp. 189–205, 2009.

[25] S. Mukherjee and K. Karan, "Human action and event recognition using a novel descriptor based on improved dense trajectories," *Multimedia Tools and Applications volume*, vol. 77, pp. 13661–13678, 2018.

[26] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1-13, 2018.

[27] A. Sargano, P. Angelov, and Z. Habib, "A Comprehensive Review on Handcrafted and Learning-Based Action Representation Approaches for Human Activity Recognition," *Appllied Science*, vol. 7, no. 1, pp. 110, 2017.

[28] A. Elboushaki, R. Hannane, K. Afdel, and L. Koutti, "MultiD-CNN : A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences," *Expert Systems with Applications*, vol. 139, pp. 112829, 2020.

[29] B. Meng, X. Liu, and X. Wang, "Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos," *Multimedia Tools and Applications*, vol. 38, no. 11, pp. 26901-26918, 2018.

[30] M. Asadi-aghbolaghi et al., "A Survey on Deep Learning Based Approaches for Action and Gesture Recognition in Image Sequences," *FG 2017 - 12th IEEE Conference on Automatic Face and Gesture Recognition*, pp. 476–483, 2017.
[31] X. Yang, P. Molchanov, and J. Kautz, "Multilayer and Multimodal Fusion of Deep Neural Networks for Video Classification," *MM '16: Proceedings of the 24th ACM international conference on Multimedia*, pp. 978–987, 2016.
[32] J. Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond Short Snippets: Deep Networks for Video Classification," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[33] N. A. Rahmad, N. A. J. Sufri, N. H. Muzamil, and M. A. As'ari, "Badminton player detection using faster region convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1330–1335, 2019.
[34] M. R. Tora and J. J. Little, "Classification of Puck Possession Events in Ice Hockey," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 147-154, 2017.
[35] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee, "Multi-label class-imbalanced action recognition in hockey videos via 3D convolutional neural networks," arXiv:1709.01421, 2018.
[36] L. Kong, D. Huang, J. Qin, and Y. Wang, "A Joint Framework for Athlete Tracking and Action Recognition in Sports Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 532–548, 2020.
[37] J. He, K., Zhang, X., Ren, S. and Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.
[38] H. Jiang, Y. Lu, and J. Xue, "Automatic soccer video event detection based on a deep neural network combined CNN and RNN," *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 490–494, 2017.
[39] Y. Hong, C. Ling, and Z. Ye, "End-to-end soccer video scene and event classification with deep transfer learning," *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp. 1–4, 2018.
[40] J. Yu, A. Lei, and Y. Hu, "Soccer Video Event Detection Based on Deep Learning," *International Conference on Multimedia Modeling - MMM 2019: MultiMedia Modeling*, *Springer*, vol. 8936, pp. 377-389, 2019.
[41] S. V. Mora and W. J. Knottenbelt, "Deep Learning for Domain-Speficis Action Recognition in Tennis," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 170–178, , 2017.
[42] M. G. Ibrahim MS, Muralidharan S, Deng Z, Vahdat A, "A hierarchical deep temporal model for group activity recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1971–1980, 2016.
[43] V. Ramanathan, J. Huang, S. Abu-el-haija, A. Gorban, K. Murphy, and L. Fei-fei, "Detecting events and key actors in multi-person videos," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[44] J. Lee, D. W. Nam, J. S. Lee, S. Moon, K. Kim, and H. Kim, "A study on composition of context-based soccer analysis system," *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 886–889, 2017.
[45] H. C. Shih, "A Survey of Content-Aware Video Analysis for Sports," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1212–1231, 2018.
[46] P. Felsen, P. Agrawal, and J. Malik, "What will Happen Next? Forecasting Player Moves in Sports Videos," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3362–3371, 2017.
[47] Z. Xu, Y. Yang, and A. G. Hauptmann, "A discriminative CNN video representation for event detection," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1798–1807, 2015.
[48] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[49] I. Rodríguez-Moreno, J. M. Martínez-Otzeta, B. Sierra, I. Rodriguez, and E. Jauregi, "Video activity recognition: State-of-the-art," *Sensors (Switzerland)*, vol. 19, no. 14, pp. 1–25, 2019.
[50] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," arXiv:1901.06032, pp. 1–62, 2019.

BIOGRAPHIES OF AUTHORS

Keerthana Rangasamy received B.E (Hons) in Biomedical Engineering from Universiti Teknologi Malaysia (UTM) in 2018. Currently, she pursuing her PhD in Biomedical Engineering under School of Biomedical Engineering and Health Science, Faculty of Engineering in Universiti Teknologi Malaysia. Her research interests are artificial inteligence, computer vision and deep learning.

Dr. Muhammad Amir Bin As'ari holds a PhD in Biomedical Engineering from Universiti Teknologi Malaysia. His Ph's work was in the field of assistive technology, computer vision and image processing and his work focused on developing a novel 3D shape descriptors for recognizing the activities of daily living based on Kinect-like depth image.He pursued his master degree and bachelor degree at Faculty of Electrical Engineering, Universiti Teknologi Malaysia, majored in Electrical Engineering. Currectly, he is working on the development of signal and image processing approaches with intelligent intervention in assistive technology and rehabilitation as well as sport performance technology especially in automated human movement recognition

Nur Azmina Rahmad received B.E (Hons) in Biomedical Engineering from Universiti Teknologi Malaysia (UTM) in 2017. Currently, she pursuing her PhD in Biomedical Engineering under School of Biomedical Engineering and Health Science, Faculty of Engineering in Universiti Teknologi Malaysia. Her research interests are artificial inteligence, computer vision and deep learning.

Nurul Fathiah Ghazali received Diploma in Electrical Engineering (Power) from Universiti Teknologi Malaysia, Kuala Lumpur (2013) and Bachelor of Engineering in Biomedical from Universiti Teknologi Malaysia, Johor Bahru (2017). She is currently working on her PhD studies in Biomedical Engineering at Universiti Teknologi Malaysia, Johor Bahru and her current research is about activity recognition study in badminton sport using inertial sensor which collaborated with Sport Innovation and Technology Centre (SITC) and Institute of Human Centered Engineering (iHumEn) from UTM JB. Her research interest includes activity recognition studies, signal processing, and machine learning.

Saharudin Ismail graduated from Bachelor of Occupational Therapy (Honours), UKM in 2008. He pursued his Master of Philosophy in Rehabilitation Technology on 2013 and attained his PHD in Health Science on 2019 from Universiti Teknologi Malaysia. Most of his work is related to Assistive Devices, Rehabilitation, Biostatistics, Motor Control, Motor Learning and Kinesiology through the application of the inertial sensors.