AUTOMATIC HIGHLIGHT GENERATION FOR SPORTS

Vishwas Shalom C [1]
Electronics and Communication
St. Joseph's College of Engineering
Chennai, India

Rubeshwaran A[2]
Electronics and Communication
St. Joseph's College of Engineering
Chennai, India

Abstract: Watching sports is an interesting way to enjoy one's day. But spending valuable time on watching matches which last multiple hours could be off putting for many individuals. Many studies and methods have been introduced to compress matches into something called highlights. Key moments or important events that occur throughout the entirety of the match can alone be viewed to get a good idea of what happens in the match. By automatically extracting these moments with the help of CNN's and OCR we can develop a model to analyse the sports video to identify key events. YOLOv3 can be used as a detection algorithm to detect scoreboards, wickets, goals etc. Along with these and with the help of moviepy highlight generation can be significantly made effective and less time consuming.

Keywords:

Optical Character Recognition (OCR), Convolutional Neural Network (CNN), You Only Look Once version3 (YOLOv3).

I. INTRODUCTION

Manual Highlight Generation is a tedious process to accomplish and labour intensive to do on the fly during a sports event. Highlights are basically moments during the game which are considered to be special or interesting to watch. During these sections of the game the audience are fascinated and leading them to cheer, applaud or boo the contenders. Using this to our advantage we can observe the audience reactions using visual deep learning techniques with the help of a linear classifier to differentiate between highlight and other parts of the game.

Recently there has been a decreased viewership of sports in live relay due to the hectic lifestyle of people. So, people of all ages have opted to view only the key moments of these games to save much valuable time. By identifying the behaviour of the audience during these events and by implementing methods such as 3D Convolutional Neural Networks and pre-trained models on various existing datasets.

Objective

The motivation of this study is to delve deeper into understanding ML and Deep Learning to make hard intensive labour easier and more automated. This paper is solely focussed on creating a solution to tackle the problem of analysing hours of video data and even more hours spent in processing the data without loss of information.

Fig 1.1 Non Highlight clip with no audience reaction

Fig 1.2 Hghlight Clip with audience reaction

Automatic Highlight generation helps editors, youtubers, streamers to automatically put out footage which is more exciting for audience to watch rather than spending hours on end watching the full video.

Another goal is to make known of features like this which can help anyone improve upon this mode in the future.

II. LITERATURE REVIEW

The primary method of highlight generation is by identifying the players' every move and action and knowledge of what happens in the game at that time. *Zhou, Y. et al* (2019) [5] modelled highlights using a finite state machine where each individual highlight is described as $G^h = (S^h, E^h)$ where S^h is set of nodes for each state and E^h is the set of edges for every event. The authors tackle the problem by using LSTM networks which can be classified as an RNN (Recurrent Neural Network).

'Other important work that we considered was of *Huang-Chia Shih (2017*) [2] who has created different methods according to various sports and the type of detection. Content-aware video analysis is a key ingredient in the method they used. By using a hierarchical content model, they detect objects in videos which are then made into events them summarized in highlights.

The work of *Zhonghan Zhao et al* (2021) [1] uses a three-step method of deep learning namely- Perception, Comprehension and decision. Similar to *Shih Huang – Chia* (2017) [2] they set up a hierarchical approach and by using already existing datasets to find out characteristics and limitations which can allow them to point out future trends in the same domain.

Huang-Chia Shih, (2017) [2] Member, IEEE has suggested the content pyramid model for a more hierarchical approach and from a spatiotemporal point of view. Also, they have used object extraction, object tracking, tactic analysis via tracking data (example of SAETA – Smart Coaching Assistant for volleyball relies on sensing the environment and the players present), camera analysis for multiple cameras in different directions.

The work of *Cheng-Yang Fu et al* (2017) [4] accomplishes an interesting method of highlight generation for online live streams by live chat reactions and comments. This is implemented by using multimodel lv-LSTM to identify if a frame in the input video is to be labelled as a 'highlight frame' or not.

Implementing YOLOv3 (You Only Look Once) allows for object detection and in our case successfully is able to detect the scoreboard according *to Bochkovskiy et al (2020)* [8] and to identify the score of the game *Liu, W., et al (2016)* [9] and *Smith R (2007)* [10] have used OCR (Optical Character Recognition) which can monitor scores from the scoreboard by finding the bounding box of the scoreboard.

The work of *Albert, J., et al (2018)* [12] explains in depth about the use machine learning for analysing a sport to identify features such as player performances using deep learning. Deep learning is a part of machine learning which is more based on artificial intelligence which learns more the more data provided to it according to *Ghanem B and Nafkha. A (2021)* [11]

III. EXISTING SYSTEM

The most commonly used main stream method of video analysis is by transcribing an input video into several events by use of Viterbi decoding algorithm in HMM which applies dynamic programming to identify the maximum livelihood path in a time event lattice. Single vector representation of a video clip has led to having a highlight clip of vector dimension $K + K^2$ (K^2 vector is an event pair) which is an efficient representation for highlight identification. The ROC (receiver operating characteristic curve) of the classifier model used in the above method using CH, CH+HOG, HOG etc is shown.

Obstacles and challenges

The most obvious challenge to a retrieving is the number of hours that game runs. To view them again and reviewing manually is impractical and time consuming. There is also the problem of bias or human error. A third error that may crop is the individual editor's point of view of editing. Added to this is the time within which such a large amount of data needs to be processed. Along with this is the high resolution that is required to maintain. These are the common factors that act as obstacles and challenges in retrieving data manually.

IV. PROPOSED MODEL

First, to deal with the drawbacks and problems faced in manual editing and the existing method, an advanced deep learning algorithm YOLOv3 is used to detect the scoreboard. So, scoreboard frames can be easily recognized

and categorized based on the high achieving sections of the event.

Secondly, the video of the game to be retrieved is exported to a directory and then using about 200 to 300 images with scoreboards with the help of the Annotation file the dataset is trained.

Followed by which *moviepy* library is used to concatenate these together to receive final highlight video.

The general workflow of the proposed model is as shown below

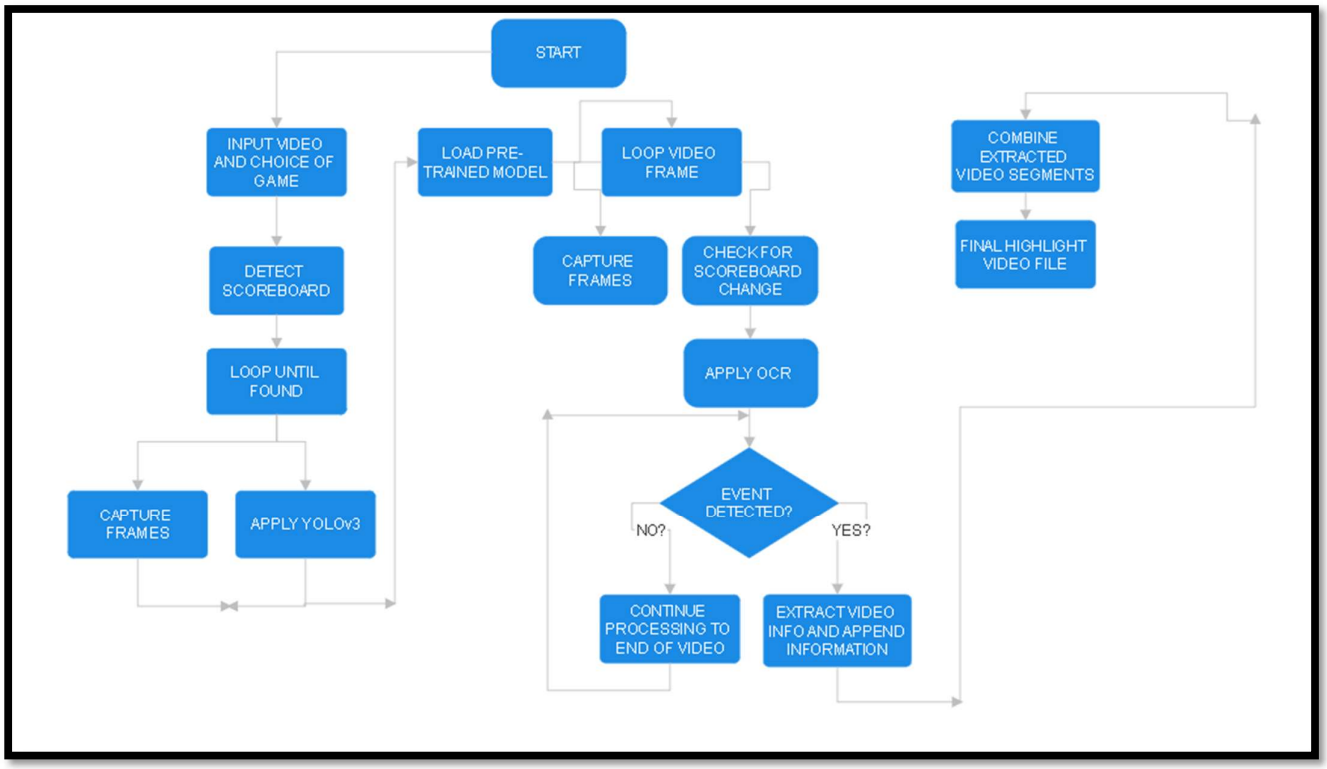

Fig 2 – General Workflow of the Proposed Model

The performance metrics is then analysed and measured and the dataset is manipulated according to requirement. Followed by this, the OCR is run on the prepared character images where each image is named according to the character. Once manipulated and validated, it is then implemented into the main code to detect events and obtain segments of highlights.

Methodology :

1. Scoreboard and Audience Detection

In any video retrieval form, there may be drawbacks in spite of the various analytical methods. Other counteract methods need to be followed to ensure that our retrieved videos are significant, precise and devoid of errors.

- To counter the drawbacks of manual editing, the implementation of video analysis using object detection can be used. This allows us to precisely identify the reaction of audience. This is possible due to Image AI's custom detection model called YOLOv3 (You Only Look Once) which is specialised in this domain.

- Custom Object Detection allows YOLO to be virtually run on any custom-made dataset. For our case, it has been implemented on the football, cricket and *kabaddi* dataset. This process of creating a dataset does initially take time but when one looks at it as the bigger picture it greatly saves time.

- From an input image (say the scoreboard) YOLO can predict the bounding box of the image grid and make probabilities of objects in the grid thus allowing us to detect multiple objects.

2. Processing the Detected Images

OCR or Object Character Recognition allows us to read the detected images of the scoreboard. Tesseract OCR is a module specifically used here which differentiates light and dark pixels to recognize the score. The corresponding part is cropped and sent through the engine.

Tesseract returns the extracted text which in our case is the score, this information is processed to determine the wickets, score etc.

3. Video Editing

After several video frames are retrieved, which may number anywhere between 40 to 200 based on the screen time, editing work must follow suit. To ensure that this work is efficiently done, a python library is used. A python library referred to as *moviepy* is commonly used for video editing, merging, resizing etc. The small sections of the video where highlights occur are segmented later and concatenated into a single final video which contains every individual key events.

As you see in the picture given below, fig 3.1, the OCR detects objects such as wicket, score, target, and the number of overs in the cricket game.

Similarly, in fig 3.2, the OCR has detected the goals, team name, and time as in the game of soccer.

V. EVALUATION OF PROPOSED MODEL

Fig 3.1 Score, Wickets, Target on OCR

Fig 3.2 Teams, Score and time on OCR

Performance Metrics

```
Found 400 images belonging to 2 classes.
Found 100 images belonging to 2 classes.
Epoch 1/10
25/25 [==============================] - 6s 225ms/step - loss: 1.0054 - acc: 0.5650 - val_loss: 0.5278 - val_acc: 0.7292
Epoch 2/10
25/25 [==============================] - 6s 227ms/step - loss: 0.5579 - acc: 0.7475 - val_loss: 0.3119 - val_acc: 0.8750
Epoch 3/10
25/25 [==============================] - 6s 224ms/step - loss: 0.4715 - acc: 0.7900 - val_loss: 0.2925 - val_acc: 0.8854
Epoch 4/10
25/25 [==============================] - 6s 223ms/step - loss: 0.4065 - acc: 0.8500 - val_loss: 0.2786 - val_acc: 0.8646
Epoch 5/10
25/25 [==============================] - 6s 224ms/step - loss: 0.4017 - acc: 0.8225 - val_loss: 0.3821 - val_acc: 0.8125
Epoch 6/10
25/25 [==============================] - 6s 224ms/step - loss: 0.3259 - acc: 0.8650 - val_loss: 0.2939 - val_acc: 0.8854
Epoch 7/10
25/25 [==============================] - 6s 224ms/step - loss: 0.3493 - acc: 0.8625 - val_loss: 0.3687 - val_acc: 0.8438
Epoch 8/10
25/25 [==============================] - 6s 230ms/step - loss: 0.3390 - acc: 0.8625 - val_loss: 0.2556 - val_acc: 0.8646
Epoch 9/10
25/25 [==============================] - 6s 230ms/step - loss: 0.2805 - acc: 0.8975 - val_loss: 0.2873 - val_acc: 0.8542
Epoch 10/10
25/25 [==============================] - 6s 232ms/step - loss: 0.2993 - acc: 0.8725 - val_loss: 0.4872 - val_acc: 0.7812
```

Fig 3.3 – Training and Validation accuracy

VI. RESULTS AND DISCUSSION

Thus, through the automatic highlight generation, any number of videos amounting various lengths, themes, and events can be retrieved and used in video platforms.

VII. CONCLUSION

Sports is a very complex but energising event which captivates the attention not only fans but also viewers across the world. Sports events take place throughout the world and all the

days of the year. Each sports event is a big achievement by itself. In order to capture the moment, highlights are used to ensure that the audience cheer and the team's performance remain fresh in the memory of the fans and viewers.

The best way to achieve the above is highlight generation. Highlight generation, as we saw in the article, is tedious by itself. We must devise automatic highlight generation to be able to efficiently generate such captivating moments for the screen and social media. To achieve our goal of automatic highlight generation the implementation python, python libraries such as keras, OpenCV, YOLOv3, OCR has allowed us to construct these user-friendly models for common use. The user then provides a video input of a sports match and its corresponding highlight clips will be generated in segments. These then are concatenated together using moviepy to output a video containing every highlight. A user-friendly design without much complex use of algorithms is retrieved. This again is important because this can efficiently and automatically generate highlights with a simple and effective output.

VIII. REFERENCES

[1] A Survey of Deep Learning in Sports Applications (2020).: Perception, Comprehension, and Decision Zhonghan Zhao, Wenhao Chai, Shengyu Hao, Wenhao Hu, Guanhong Wang, Shidong Cao, Mingli Song, Senior Member, IEEE, Jenq-Neng Hwang, Fellow, IEEE, Gaoang Wang Member, IEEE. Automatic Sports Highlights Generation: A Survey. arXiv preprint arXiv : 2012.10238.

[2] Shih, Huang-Chia. "A Survey on Content-aware Video Analysis for Sports." IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2017).

[3] Gade, R., & Moeslund, T. B. (2014). Automatic highlight detection in sports: A review. IEEE Transactions on Multimedia, 16(6), 1610-1625.

[4] Cheng-Yang Fu, Joon Lee, Mohit Bansal, and Alexander Berg. (2017).Video Highlight Prediction Using Audience Chat Reactions. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 972–978, Copenhagen, Denmark. Association for Computational Linguistics.

[5] Zhou, Y., Liu, C., Cui, Y., Wang, X., & Yang, S. (2019). Real-time Sports Highlights Generation with LSTM Networks. In 2019 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1574-1579). IEEE.

[6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[7] Assfalg, Jrgen, et al. ”Semantic annotation of soccer videos: automatic highlights identification.” Computer Vision and Image Understanding 92.2 (2003): 285-305

[8] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

[9] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (ECCV). ArXiv:1512.02325

[10] Tesseract OCR: Smith, R. (2007). An overview of the Tesseract OCR engine. Ninth Inger International Publishing.

International Conference on Document Analysis and Recognition (ICDAR). 10.1109/ICAR.2007.4376991, Curitiba, Brazil,

[11] Ghanem, B., & Nafkha, A. (2021). Deep Learning in Automatic Sports Highlights Generation: A Survey. IEEE Access, 9, 94417-94431.

[12] Machine Learning and Data Mining for Sports Analytics 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD (2018), Dublin, Ireland, September 10, 2018, Proceedings, Albert, J., Koning, R. H., & Kok, A. A. Sports Data Mining and Analytics: Using Neural Networks to Analyse Football Player Performance. Springer International PU