Cricket Highlight Generation Using Machine Learning And Deep Learning Techniques

A thesis submitted in partial fulfilment of the requirements for the degree of B.Tech. in Computer Science and Engineering

by

Students : Prasoon Mishra (116CS0024)

Vibhor Kumar Srivastava (116CS0032)

Amit Sharma (116CS0034)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,

DESIGN AND MANUFACTURING, KURNOOL

May 2020

Certificate

It is certified that the work contained in this thesis entitled "Cricket Highlight Generation Using Machine Learning And Deep Learning Techniques" by "Prasoon Mishra (116CS0024), Vibhor Kumar Srivastava (116CS0032) and Amit Sharma (116CS0034) " has been carried out under my supervision and that it has not been submitted elsewhere for a degree.

May 2020

Dr. Sanjaya Kumar Panda
Department of CSE
IIITDM KURNOOL

Abstract

This project is aimed in producing cricket highlights with an easy and less labour-intensive approach. we propose a model capable of automatically generating cricket highlights using Computer Vision and Machine Learning Techniques.

Cricket is a large duration game with many complex rules involved. Keeping the rules in mind and the amount the time the game takes to complete we are proposing a model which can handle events like these along with event-based and excitement based features to identify and trim important events from the full length cricket match into a 7-8 minute highlight. Some of the identifying features are the audio intensity, celebration, the scoreboard, the stadium scenario etc.To identify the highlight authenticity we performed experiments like comparing our highlight with the official released ones, manually checking for the error in the highlight, checking if the features we considered are working together.

Acknowledgements

We would extend our sincerest gratitude to our project supervisor and mentor Dr. Sanjaya Kumar Panda (Department of CSE, IIITDM Kurnool) for his esteemed guidance, invaluable inputs during our project tenure. We could not have imagined having a better advisor and mentor for pursuing this project.

We choose this special space to gratefully acknowledge the immense cooperation and contribution of our honourable director sir PROF. D.V.L.N. SOMAYAJULU and DR. RENJITH P. sir, who have made this journey a truly learning and joyful experience. Their efforts in guiding us in every aspect to provide us their valuable feedback and support are highly appreciated. Not only they guided us for the successful completion of our project but also showed us ways for its future work.

We would also like to thank all our friends who helped us by providing their inputs in our project.

Contents

List of Figures

List of Tables

Abbreviations

OCR **O**ptical **C**haracter **R**ecognition

Dedicated to
Our institute IIITDM Kurnool

Chapter 1

Introduction

1.1 Overview and Problem Statement

Cricket Highlight generation is a lot of laborious work and we aim to generate a model which performs the automatic highlight generation of the cricket match video which is a challenging problem itself.There are many features in a video frame to look into like the excitement via audio as well as video (Cheering image), the scoreboard which changes after every ball.So by targeting these features we focus the highlight generation based on event-driven and on audio-driven features..

1.1.1 What is Highlight Generation

Highlight Generation is the process of trimming the most exciting and compelling events in clips extracted from the full-length sports video, this process of transforming a larger video into a smaller one involving all important and major events is also called **Video Summarization**. A full-length match can go from 3hrs or above but the final highlight should be short and within limit 7-8 minutes at a maximum.This highlight must focus on events like fours, sixes, wickets fallen, centuries/milestones achieved by any player and all other major events.

1.1.2 Types of Events and Information Obtained

First of all a full length match video contains replay and non replay parts, a highlight never contains the replays, they contains the parts of the non replays which is the actual match

FIGURE 1.1: Event Based Frame

we need.In the non replay there are many events happening audience cheering(not audio), score board changing every ball, number of balls incrementing every, players achieving milestones like centuries.

With the help of OCR score board reader we can keep track of the score changing every ball, if the score changes by 4 then the current video frame is part of a Four and the video frames nearby this frame should be marked as an important, similarly happens if the score changes by 6. score board also shows wickets fallen so we can also use this information to trim the nearby video frames as important and these to the highlight.

With the help of audience cheering (Image) we can know if any milestone has been reached, if such event is triggered we search in the nearby frames for any information regarding the reason for the cheer, for example : batsman made a century or a ball-er had a golden over etc.

1.1.3 Types Of Features

1.1.3.1 Event Based

Event based features are those features which are visible in the video like cheering of audience, visibility of score board.This involves the occurring of important events like fours, sixes, wickets fallen. These events can be captures by keeping an eye on the scoreboard and report when it changes and also the amount of change(Like from score (64-1 to 68-1: means its a four if the audio also agrees i.e. excitement of the audio). Keeping track of the each frame of the video and start marking the frame important if it ranks as significant.

FIGURE 1.2: Audio Based Frame

1.1.3.2 Audio Based

Audio-based features like audio intensity which is an important feature to take into consideration when considering a frame of video as an important. Some events like centuries achieved by a player and other major milestones can be identified with the help of audio-based features. This model keeps into the track the loudness of the audio and mark important episodes. The Figure 1.2 represents a brief about what happens simply analyzing the audio or video alone are not enough to get a proper highlight, both have to interact properly to get a proper classifier(Important v/s Non-Important video frame).

1.2 Motivation

There is a large number of cricket matches throughout the year with a lot more fans and it is very tedious and time consuming for a cricket maniac to keep track of every match, so highlights serves a great purpose in fulfilling the needs of a fan by giving the summary of the important events happened in the match within minutes. Today making a sports highlight is very laborious work, requires high analytical skilled people, a lot of manual work and it is very time consuming process too. All these factors fuels the need for creating a model which does exactly the thing and generating the highlight with minimal or no human involvement and also reducing the time needed to produce the highlight. A brief idea about the work is summarized in the Figure 1.3.

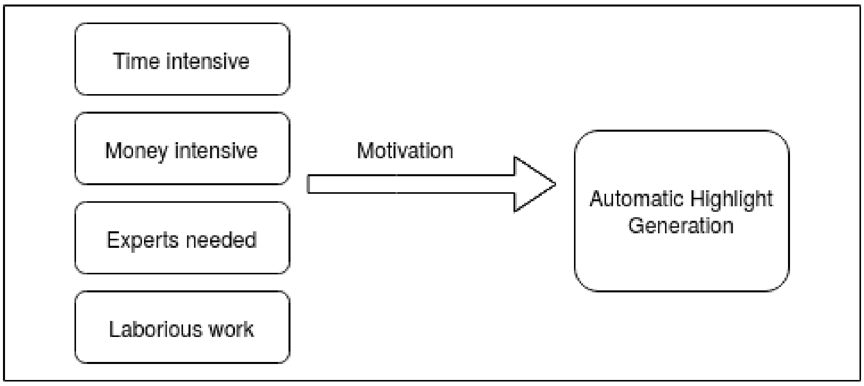

FIGURE 1.3: Motivation

1.3 Contributions of this work

Each video is made up of sequence of frames/shots and these frames are an important aspect of the cricket video summarization. These sequence of frames or video shots are part of the same match video. An example of the sequence of frames and how it is described is illustrated in Figure 1.4 we take the video frames as the building block of our model for recognition of important video frames, along with a helping hand of the audio.The reason we chose this structure for generating the highlight is :

- Each video frame are part of the entire match and they are a contributing factor into classifying frames as important/non-important.
- Each video frame contains features which cannot be obtained from the audio alone, like the scoreboard, the orientation of the camera, Cheering of the audience, identifying non-replay from replay video frames.
- looking at the orientation of the camera we can identify the start of the important event

Just marking any frame as important is not enough we have to capture the entire duration of the video clip during which that important event occurred and here the direction of camera is crucial.for example in Figure 1.5 the start of the important event can be a bowler starting to bowl and the camera angle is exactly at this position every time. We also have to throw out the unimportant content of the entire match which includes advertisements and replays. So to remove them we first have to recognize them.Model to classify the frame as a replay/(non-replay + advertisement) classifier is required. Then a Model to keep a track of the score board and mark any changes and if there are any significant

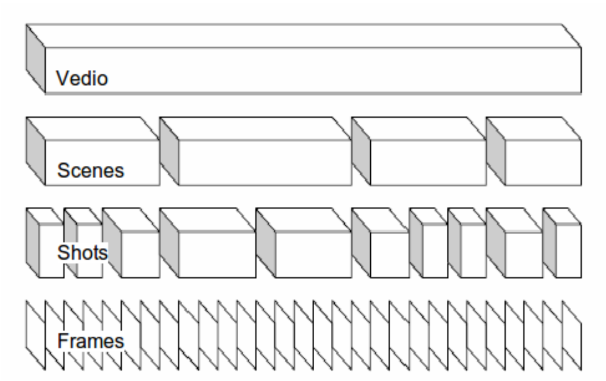

FIGURE 1.4: Frames and Their Description

FIGURE 1.5: Deciding Starting Frame of an Important Event

changes mark those video frames as important. Another Model is required to keep track of the playing field layout.i.e. keeping track of the players arrangement in it.

After marking each frame as important and non important we can take into consideration the nearby frames to identify the start of a particular important event and then also the end of it. When this is decided we trim those clips and merge them into a single highlight. For example the entire process is in the Figure 1.6.

Existing works include generating highlight based on audio which, input the full match video, extract the audio, break the audio into chunks, compute short-time energy of every chunk, classify every chunk as excitement or not (based on a threshold value), merge all the excitement-clips to form the video highlights. Considering the audio as the only feature then there will be more errors for example intensity of the audio will be high during a four, sixes, wickets but it will also be high during cheering for a particular player, during the advertisements, commentary of the commentators.so this approach will not be effective alone it requires other factors as a added feature too.

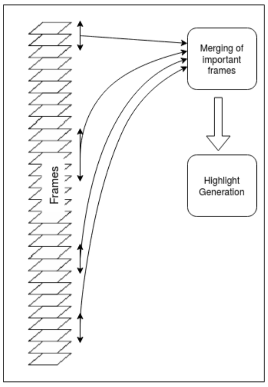

FIGURE 1.6: Entire Video Frame Importance Detection Process

Our work is based on approaching the problem with using the event driven model with a bit of a helping hand of audio driven model which is a more promising model because now we can make sure only the important video frames are taken as the part of the highlight. Varying parameter measured : the video frame number, the importance assigned with each video frame, replay/non-replay label to each video frame, excitement intensity high/low of current video frame, whether video frame is part of Four/Six/Wicket/Milestone category, estimated duration of time each ball from delivery to strike/miss takes, starting frame No. and ending frame No of an important event.

There are important features that needs to be considered for example after marking the video frame as important how further and backward need to go around that video frame to get the entire important video clip.This is done experimentally and we realized that when the score changes(say for example difference by Four) then going **¡x¿** frames back where the camera is above the bowler when he/she starts bowling marks the start of the important event and this important clip ends after **¡x¿** frames when the bowl approximately reaches the boundary.for example the Figure 1.7 shows the way the important clip is trimmed using the important video frames from the full match video.

1.4 Organization of the thesis

The report consists of the following chapters:

- Chapter 1 – Introduction
- Chapter 2 – Background and Related Work
- Chapter 3 – Proposed Mechanism

FIGURE 1.7: Marking the Beginning and End of the Relevant Clip

- Chapter 4 – Performance Study of system
- Chapter 5 – Conclusions and Future Work

Chapter 2

Background and Related Work

2.1 Background

While generating highlight of a cricket match some factors which remain consistent are: the camera angle and range, the score board position in the video, duration of a ball delivery to strike/miss by batsman. As per the ICC Standard Playing Conditions 2015–16 for international matches:

- In the Figure 2.1 Where matches are broadcast the camera specification set out as per following shall be mandatory as a minimum requirement.

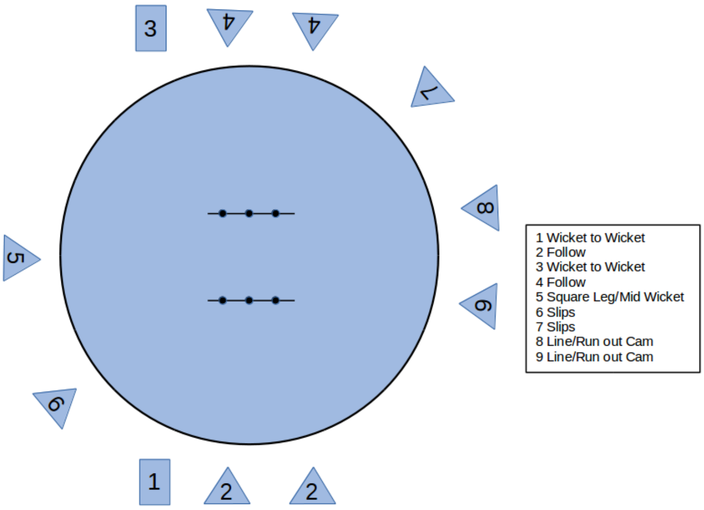

FIGURE 2.1: Minimum Camera Angle Specifications Where Matches are Broadcast

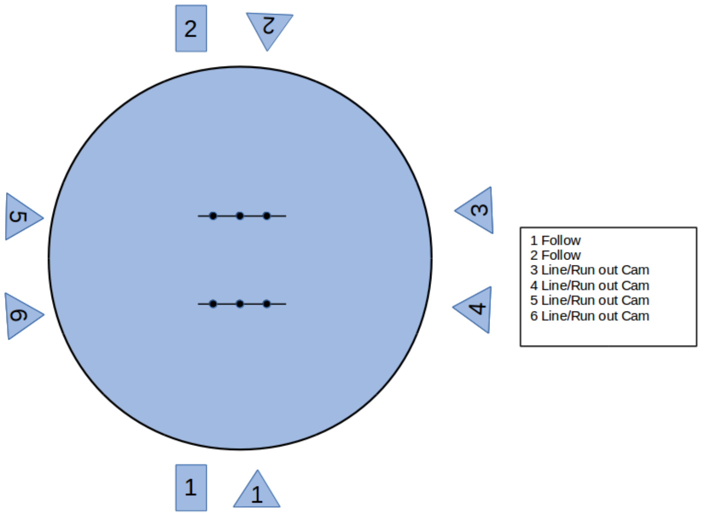

FIGURE 2.2: Minimum Camera Angle Specifications Where Matches are Not Broadcast

- In the Figure 2.2 Where matches are not broadcast the camera specifications set out as per following shall be mandatory as a minimum requirement.

2.2 Related Work

According to **Maheshkumar ET. AL.**[2], highlights can be generated by obtaining some information from audio channels by estimating audio energy and short-time zero-crossing rate which registers excitement levels. Further for getting current score a user defined particular rectangular is analysed. Before passing the video to the system to generate the highlight, all the advertisement frames are removed by manual video editing or tools as a preprocessing step. This preprocessing is required because there are chances of an advertisement frame with high sound to be included in the highlight. The average precision observed during performance study was 85.32 % and 89.74 % for two different videos.

Daniel ET. AL.[3] proposed a significant approach of Bowler Run-Up Sequence for cricket highlight generation. This paper makes use of Oriented-Fast Rotated Brief (ORB) method for Bowler Run-Up Sequence (BRS) .The ORB method is suited for BRS since it reduces the computational complexity and it is possible to generate real time highlights using this method. The method proposed detects BRS correctly with precision of 99.1 %. The limitations of this paper were -

- Only the BRS in a cricket video was detected but highlight generation requires detection of other events also.

- Each time to generate highlights of a cricket inning, the model needs to be trained because the difference in clothing of players can lead to bad detection.

Maheshkumar ET. AL.[4], proposed a different approach of highlight generation which uses two levels of abstractions- events and semantic concepts. In this paper, the game specific concept selection and event selection criterias have been presented. A degree of abstraction parameter has been used which decides how shortly the concepts can be produced in the final highlight. One of the limitations of this work also was that, in this proposed model requires a preprocessed video removing all the advertisement frames which is time consuming.

Maheshkumar ET. AL.[5], proposed a Dynamic Programming based on the Hidden Markov Model (HMM-DP) approach for structure analysis of cricket video sequence. A scaling factor is introduced in the dynamic programming recursions to improve robustness in shot detection. 11 different cricket video sequences were used to train the Hidden Markov Model then to generate the highlight of unknown video the likelihood sequence for each frame is calculated and the frame is classified as one of the 11 classes. During the performance measure, 87.5 % overall classification for motion detection and 75 % for color likelihood were observed with scaling factor (alpha=2) in the dynamic programming recursions. It also requires lot of training time and space to train the models.

Somnath ET. AL.[6] proposed an approach where a sports video is hierarchically divided into temporal partitions namely, megaslots, slots, and semantic entities, namely concepts, and events. For the highlight generation, the concepts and and events are selected from all frames and then concatenated in order to generate the highlight. Further a degree of abstraction(n) was introduced as a controlling parameter to revise the highlight video. This approach also required manual preprocessing of original video before passing to the highlight generator system for highlight generation.

2.3 Conclusion based on literature review

On the basis of following we have inferred that the existing designs of automatic cricket highlight generation require professional expertise or some of them which are based on audio based approach of highlight generation still require improvement. So for the design of our project we have listed down some of the major issues that need to be solved for better cricket highlights.

- Manual highlight generation requires professional skills and is time consuming.
- Audio based highlights can misclassify the frames, as all the video frames with high sound energy may not be part of the highlight but are included.
- Not all exciting events are cheered by commentators and spectators, so they may infer that the particular frame is not part of the highlight.
- Chances of advertisements which have a high intensity of audio, to be included in highlights will be more which is irrelevant.

Chapter 3

Proposed Mechanism

3.1 Necessary Network/System Architecture

Machine learning, computer vision and NLP based techniques are widely used in other domains but they have not been used much for highlight generation which is a kind of video summarization. In our design we have tried to overcome this issue. Our system consists of neural networks and machine learning models which gives better results than existing audio based systems in different parts of the highlight generation phase such as replay/no-replay classification, scoreboard digit recognition, bowler run-up detection etc. So, to build such a system a computer with sufficient memory and computation power is required.

3.1.1 Highlight Generation Workflow

In all the approaches of highlight generation, prepossessing of the original video was required before passing it to the system. In our proposed approach we have tried to remove the prepossessing step since we wanted to remove the manual editing in highlight generation. The input video is directly passed to the system without any prepossessing and frame by frame analysis is done. Since it takes much time to analyze frame, we get the video shot (collection of frames taken from same camera angle) and pass a single key frame from this video shot to next level and if that frame can be part of highlight, we add the entire video shot into the highlight otherwise don't add the video shot into the final highlight. n

3.2 Problem definition

Trivially highlight generation takes lot of work and skills in graphic editing software, but our intention of achieving is to reduce the manpower and skills required to complete the same task.Cricket Highlight Generation Using Machine Learning And Deep Learning is a idea which generates a highlight of any cricket match video by taking as input the match video and outputting the 5-8 minutes highlight video which covers all major and important events in the match without repetitions.

Many influencing factors are ignored when trying to arrive at a solution and we eventually end up in negative solution or even end up in more problematic outcome.so one should know the obstacles beforehand and understand them well so that they can be handled better now and neutralized later. The challenges involved during the development of this model were for example How to recognize whether the current frame is that of a replay or a non-replay, Another challenge is finding more other factors which can effect the highlight and give a better result if used.

However this problem is a highly complex one and falls under **NP-hard** problem, more or less our aim of the model is check whether the video frame is important or not, which is similar to knapsack problem where we have to decide whether to put the item in knapsack or not, this leads to the problem being complex and **NP-hard**.

3.3 Proposed Idea

The idea and the proposed model is described in the Figure 3.2.The whole highlight generation model is divided into 3 stages First the Full Match is divided into video snapshots which are specific moments in match, then these video shots are given to the next stage which analyzes those snapshots, analyzing the snapshots involves detecting the scoreboard in the snapshots and looking for desirable changes like Fours, Sixes, Wickets.

Analyzing the snapshots also involves discarding those video snapshots which are part of a replay and keeping the other ones, if the current snapshot shows any useful changes then this video snapshot is marks as the important and will be used in generating the highlight.A start-checkpoint is marked at the start of the important event which involved detecting the frame where bowler starting his run-up and the end-checkpoint is marked to the video-shot where the ball hits the boundary.

1. Take the complete video match, Frame by Frame :
    a. Replay ?= Check if the current frame is Replay or Non-Replay
    b. If (Replay)
        i. Continue
    c. Else
        i. Detect scoreboard digits from scorecard using OCR technique.
        ii. Compare detected score from previously stored score.
        iii. If (score changed is more than 4 or 6 or change in wicket column)
            1. Go to previous few frames(k), where bowler starts bowling and from that frame to next few frames continue adding frames to the highlight Video.
        iv. Else if (Excitement threshold('t') is achieved)
            repeat step c-iii
    d. Update previous scoreboard with current scoreboard if frames have been added to the highlight or leave the score as it is.

FIGURE 3.1: Pseudo Code

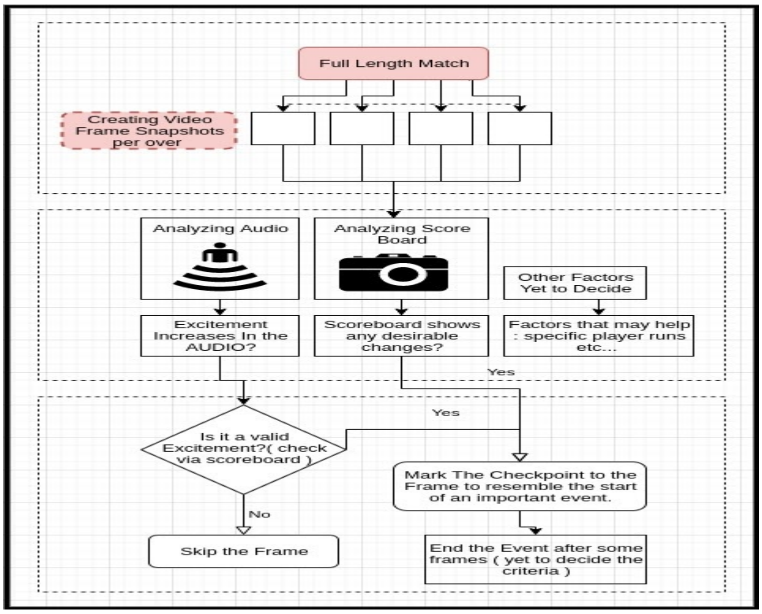

FIGURE 3.2: Proposed Model

When the frames are marked as important and the start-checkpoint and end-checkpoint are marked in the entire video. The individual start-end checkpoints are trimmed out and put in to the actual highlight video of the match. The Figure 3.1 shows the pseudo code used for developing the model.

3.4 Implementation Details

There are different models used for specific parts of the bigger highlight generation framework, all them models and ideas used and their working are mentioned below:

FIGURE 3.3: start-checkpoint is when the bowler runs to deliver the ball, end-checkpoint is when the ball reaches the boundary or umpire shows the hand sign for out.

3.4.1 Dividing Entire Match Video into Video Snap Shots

Each match video is divided into overs and we extract 42 video snapshots from each over. the idea behind using video snapshots instead of frames is that, the entire match can be of hours to days and it not computationally feasible to use frame by frame analysis approach instead using some specific frames as the video snapshots which describe change in frame, i.e if two frame have similarity greater than 80% then they are almost the nearby frames and can be discarded as it will lead to repetition of knowledge. Similarity is measured via SSIM and then these snapshots will not be used directly in the highlight, if a video snapshot is marked as important i.e. it shows significant change in the scoreboard then we move back from that snapshot to the one where the bowler is running to deliver the ball and mark it as the start-checkpoint and then move forward from that snapshot to the one where the ball hits the boundary and umpire shows the hand sign and mark it as end-checkpoint. The Figure 3.3 shows the frame which is used as the starting and ending checkpoints. Then this start-end checkpoints are referred from the actual video and trimmed that particular part and added to the highlight video.

3.4.2 Replay and Non-Replay Classification

A match involves many other events going on aside from the actual important ones, like the replays and then the people cheering, the advertisements and we need only the events snapshots which occurred during the non-replay where the actual match is happening and discard other unimportant video snapshots. The idea is whenever there is a advertisements or a replay the scoreboard which is always present at the bottom left corner is not visible and hence we can declare the video shot as a replay.

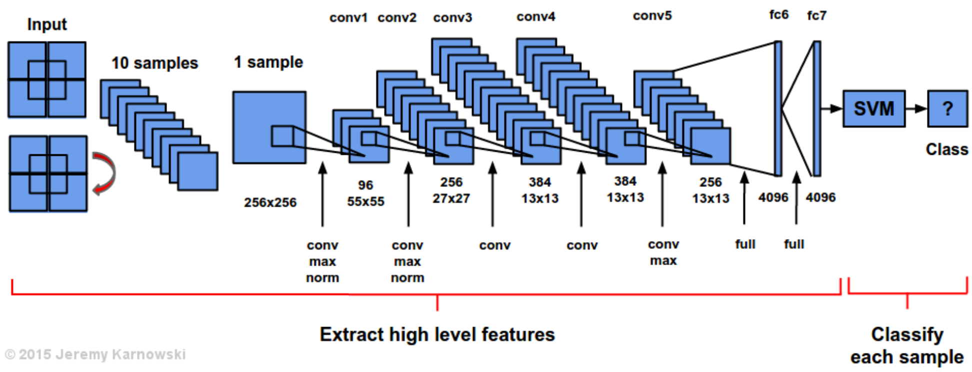

FIGURE 3.4: Alexnet+SVM Model [1]

(a) Non-Replay: Notice Scoreboard is present (b) Replay: Notice no Scoreboard is present

FIGURE 3.5: Identifying Non-Replay from Video, If the scoreboard is present its a Non-Replay otherwise it is a Replay/Advertisements

This is achieved by cropping the lower left part of the video snapshot where there should be scoreboard and pass it through the trained **Alexnet + SVM** model.Figure 3.4 shows the architecture of the Alexnet + SVM model. In this model the **AlexNet** outputs a 4096 1-D vector which is a representation of the cropped image and is then passed to the trained Binary Support Vector machine which classifies the image as being Replay or Non-Replay. The Figure 3.5 shows the idea behind classifying a video snapshot as a Non-Replay or a Replay+Advertisements.

3.4.3 Scoreboard Reader and Detection

Scoreboard keeps track of the ball-by-ball enumeration on the video during a match. Scoreboard is describes some of the major events and used to predicts events like Fours, Sixes, Wicket. Scoreboard shows current score and number of wickets fallen, and it also

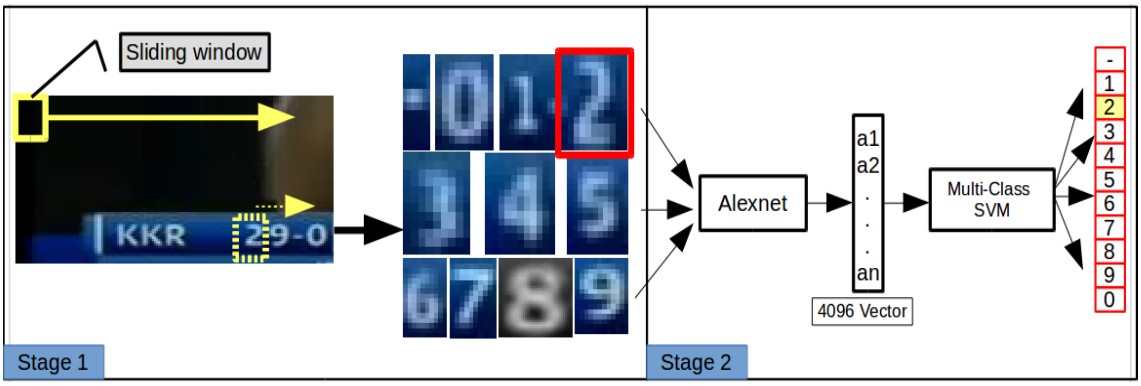

FIGURE 3.6: 2 Stage Sliding Window Model to Predict The Score From The Scoreboard

shows other statistics occasionally like current over, speed of delivery, batsman current runs.

So it is crucial for a model to accurately perform the detection work to keep the score and other features in check as these features changes quickly during the match.Scoreboard Detection and Reading can be performed by using a Optical Character Recognition(OCR) model.

There are two stages to developing a OCR, **First** a Digit recognition model is designed using Alexnet + SVM model. Here the Alexnet generates a 1-D 4096 vector which is representation of each symbol and then it is passed to a trained Multi Class Support Vector Machine which classifies the image to one among the 11 symbols.The 11 symbols are 0 to 9 and "-".Dash symbol is used to separate the score from the number of wickets fallen. **Second** a sliding window model which keeps moving over the image. The part of the image,where window is on is cropped and passed to the CNN+SVM to predict the symbol. This action is performed over the video shot using sliding window and thus we get the score and wickets fallen in a array. The Figure 3.6 shows the 2 stage sliding window model to predict the score on the scoreboard.

Chapter 4

Performance Study of system

4.1 Implementation/Simulation Environment

For the simulation purpose, all the implementation code was written in Matlab software and for testing of small submodule jupyter notebook was used to test submodule functionality in python. We used matlab for our implementation because it provides easy syntax for writing machine learning and deep learning models, i.e. for implementating a neural network just a function call is required. It saves time comparison to other software packages where scratch implementation is required. Our objective was to build the system rather than impelementing just the model, so in matlab we directly used the prebuild function for implementing the neural network model.

4.1.1 Evaluation

Highlight generation is a subjective thing. Thus there is no ground truth. Hence we adopted the strategy of finding intersection of union metric to capture the overlap of each frame and find out the mean intersection of union. This is generally a good metric for overlap and is widely used in object detection benchmarks.

4.2 Results and Analysis

We compared our generated highlights with the official highlights to find how accurate our result have been. We used intersection of union metric to capture the overlap between

Table 4.1: Comparison of highlights generated by different models

$Model$	$MeanIOU$
Audio based approach	0.53
$CNN + SVM$	0.61

both the highlights frame by frame. We were able to achieve mean IOU of 0.61. It shows that there is a significant overlap between our highlight and the official highlight generated but still there are some difference in them.

We compared our highlight with the highlight generated by audio based method and it was significantly better.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

We started this project to propose a system which can generate the highlight of a cricket match using machine learning and deep learning techniques. We started our design by following audio based approaches and later we used image processing and neural networks to generate the highlight. Our goal was to automate the highlight generation process which we accomplished upto larger extent.The results obtained were better than audio based approach highlight generators but still improvements are required to achieve the better accuracy.

5.1.1 Limitations

There are few limitations to our work, as our proposed approach is slightly longer and has error compared to official highlight. We also could not compare our result with official highlight in metrices form because chances of false positive t be included are more in our proposed system. Sometimes event based highlights and audio based highlights overlap in few frames, which need to be included only once in final highlight, these limitations can be removed in future works.

5.2 Future work

In future, this work can be extended to include addition functionality such as :

- This system can be designed to generate highlight of individual player's innings also.
- Include unusual events that took place in match into the highlight (important catch drop, injuries etc.)
- Including singles and doubles also, if important (for example final winning run).
- Later this work can be used to generate highlight of some other sports video also with modification.

Bibliography

[1] J. KARNOWSKI. Alexnet visualization. [Online]. Available: https://jeremykarnowski.wordpress.com/2015/07/15/alexnet-visualization/

[2] M. H. Kolekar and S. Sengupta, "Caption content analysis based automated cricket highlight generation," in *National Communications Conference (NCC), Mumbai*, 2008, pp. 461–465.

[3] D. Ringis and A. Pooransingh, "Automated highlight generation from cricket broadcasts using orb," 08 2015, pp. 58–63.

[4] M. H. Kolekar and S. Sengupta, "Event-importance based customized and automatic cricket highlight generation," in *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 1617–1620.

[5] M. H. Kolekar and S. Sengupta, "Hidden markov model based structuring of cricket video sequences using motion and color features," in *ICVGIP*, 2004.

[6] ——, "Semantic concept mining in cricket videos for automated highlight generation," *Multimedia Tools and Applications*, vol. 47, pp. 545–579, 2009.

[7] P. Shukla, H. Sadana, A. Bansal, D. Verma, C. Elmadjian, B. Raman, and M. Turk, "Automatic cricket highlight generation using event-driven and excitement-based features," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.