Journal of Engineering Sciences

Vol 12, Issue 06, JUNE /2021
ISSN NO:0377-9254

UMPIRE POSE DETECTION

Mr K.NARSIMHULU[1], G.AJITH,[2] J.N.ABHINAV [2], C.SANKEERTH[2]

[1]Assistant Professor, Department of CSE, Sreyas Institute of Engineering and Technology. [2]Final year B.Tech. Students, Department of CSE, Sreyas Institute of Engineering and Technology Hyderabad, Telangana, India

ABSTRACT

In recent years, there has been increased interest in video summarization and automatic sports highlights generation. In this work, we introduce a new dataset, called SNOW, for umpire pose detection in the game of cricket. The proposed dataset is evaluated as a preliminary aid for developing systems to automatically generate cricket highlights. In cricket, the umpire has the authority to make important decisions about events on the field. The umpire signals important events using unique hand signals and gestures. We identify four such events for classification namely SIX, NO BALL, OUT and WIDE based on detecting the pose of the umpire from the frames of a cricket video. Pre-trained convolutional neural networks such as Inception V3 and VGG19 networks are selected as primary candidates for feature extraction. The results are obtained using a linear SVM classifier. The highest classification performance was achieved for the SVM trained on features extracted from the VGG19 network. The preliminary results suggest that the proposed system is an effective solution for the application of cricket highlights generation.

1 . INTRODUCTION

Automatic video summarization has gained increased attention in the recent past. Sports highlights generation, movie trailer generation, automatic headlines generation for news are some examples of video summarization. The focus of the present work is sports video summarization in the form of highlights. The highlights of a game provide the summary of important events of that game such as a goal in soccer or a wicket in cricket. It is a challenging task to summarize the highlights from sports videos as these videos are unscripted in nature. An efficient approach can be based on identifying key events from the sports video and use them to automatically generate the highlights.

1.1. SCOPE

Among sports, cricket is the most popular game in the world after soccer and has the highest viewership rating. In the game of cricket, the umpire is the person with the authority to make important decisions about events on the field. The umpire signals these events using hand signals, poses and gestures. This innate characteristic of the cricket video can be leveraged as one approach for solving the problem of cricket highlight generation. Therefore, a system can be developed to detect the unique signals and poses shown by the umpire to automatically generate cricket highlights.

1.2. METHOD

We have used a method for umpire pose detection for generating cricket highlights based on transfer learning is proposed in this

work. We explore the use of features extracted from the pre- trained networks such as Inception V3 and VGG19 networks pre-trained on ImageNet dataset. A linear support vector machine (SVM) classifier is trained on the extracted features for detecting the pose of the umpire. A new dataset, SNOW, is introduced in this work and all experiments are performed on this dataset. The system built using this dataset is evaluated on cricket videos for highlights generation

2. PROBLEM STATEMENT

To generate a cricket match highlights all the work is manually done by some person, this takes a lot of time and effort .The person must be more concentrated towards the work he has to sit on it has to go through the video keenly and then he has to cut the important parts of the video and then he has to paste it somewhere else and he has to continue this work until the end of the match video . After this work he has to combine all the cut videos and has to make a video again.

DISADVANTAGES

1. Takes a lot of time

2. Takes a lot of memory

3. Should cut the important Aspects of the match into parts manually

4. The person must be paid

3. Proposed Method

A method for umpire pose detection for generating cricket highlights based on transfer learning is proposed in this work. We explore the use of features extracted from the pre-trained networks such as Inception V3 and VGG19 networks pre-trained on ImageNet dataset. A linear support vector machine (SVM) classifier is trained on the extracted features for detecting the pose of the umpire. A new dataset, SNOW, is introduced in this work and all experiments are performed on this dataset. The system built using this dataset is evaluated on cricket videos for highlights generation

ADVANTAGES

1. Where User can easily generate highlights of the required cricket mach

2. Easy to use

3. Safe Environment

4. 24/7 Availability

4. Implementation

4.1. DATASET DESCRIPTION

We have collected images of umpires performing various actions pertaining to events such as “Six”, “No Ball”, “Out” and “Wide”. These images have been obtained from various cricket match videos from YouTube and Google images. The dataset comprises of five classes of data. Four classes belonging to the four actions and one no action class in which the umpire does not perform any action.

Each class consists of 78 images summing up to a total of 390 images for all five classes.

Umpire

390 Images

• 5 Classes – Six, No Ball, Out, Wide (SNOW), No Action

• Each class consists of 78 Images

• Contains Images of Umpires with Different Colour uniforms, camera angles, lighting, etc.

Non- Umpire

• 390 Images

• Contains Images of Team players, Field, Audience, etc.

Fig. illustrates some of the images in the dataset for the four classes of events. While this dataset is sufficient for umpire pose detection, there is a need for a non-umpire dataset to distinguish between a frame containing the image of an umpire and that which does not contain an umpire. This nonumpire set contains images of team players, the playing field, crowd, etc. This will facilitate the building of classification system that can be applied directly on the frames of cricket videos

Fig 4.1:-Six

Fig 4.2:- No Ball

15

out

Wide

Fig. 3 shows some examples of images belonging to the no action class and non-umpire class. There is a lot of diversity in the images for every class. There are images shot at different camera angles, orientations and lighting circumstances. Umpire images with different color uniforms, and varying background conditions have been included in this dataset to cover a wide range of game settings. Also, successive frames with slight differences between each other have been extracted from cricket videos and included in the dataset. This dataset has been made available online and can be downloaded

Fig 4.5:- Umpires- No Action

Fig 4.6:- Umpires- Non- Umpire Class

4.2. Support Vector Machine (SVM)

To recognize the daily life activities and predict user behavior by using a decision fusion of four individual Support Vector Machine (SVM) kernel functions, where each kernel is designed to learn the performed activities in parallel.

- Fusion of four kernel functions
- Learns and performs in parallel
- Outputs a hyperplane
- Used for classification and regression analysis

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate ndimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

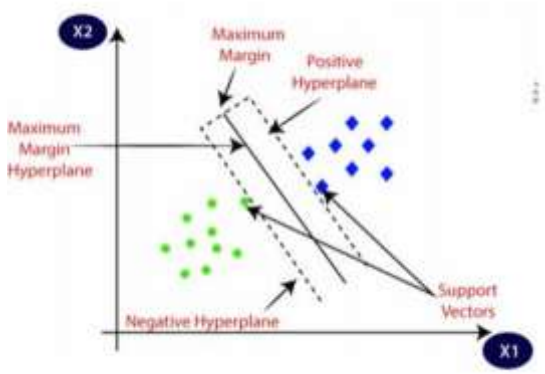

Fig 4.7:-Support Vector Machine

Example: SVM algorithm can be used for Face detection, image classification, text categorization, etc. SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm.

We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. Based on the support vectors, it will classify it as a cat. Consider the below diagram

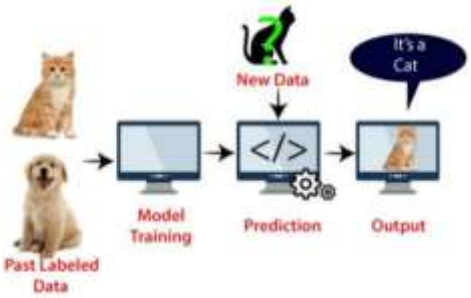

Fig 4.8:-Support Vector Machine

SVM can be of two types:

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier

5. METHODOLOGY

All experiments were performed on Inception V3 and VGG19 pre-trained networks provided in the Keras package for Python. Both networks have been trained on a subset of the Image Net dataset. The Image Net dataset is a benchmark dataset for object recognition and was used in Image Net Large Scale Visual Recognition Challenge (ILSVRC). They are trained on more than one million images belonging to 1000 object categories. Therefore, the networks have learned feature representations for a wide variety of categories in the source domain. The Inception V3 consists of 159 layers in total. The VGG19 consists of 26 layers in total with 16 convolution layers, and 3 fully connected layers.

The proposed system is built in two phases. The first phase involves designing classifiers to distinguish images containing an umpire versus no umpire, and also detect the pose of the umpire, if present. This phase involves the following steps: preprocessing, Inception V3 network and 224 by 224 pixels for the VGG19 network

The features are extracted from different layers of the pre-trained networks. Finally, these features are used to train a linear SVM classifier to output the class label of the predicted pose of the umpire. The trained classifier models are saved using the python pickle library to be used in the next phase of video summarization. The second phase involves detecting the events from the cricket videos using the saved classifier models and generating the summary of the videos. The steps involved in video summarization are: pre-processing, feature extraction, umpire detection, event detection, frame accumulation and video summary generation. The classifier design and the combined pipeline for video summarization are detailed in the following sections.

5.1. Phase 1 – Classifier Design For cricket video summarization two classifiers are designed: Classifier 1 and Classifier 2. The Classifier 1 is designed to distinguish between a frame containing an umpire versus a frame that does not contain an umpire. Two sets of data were created. One set containing all 390 umpire images from the SNOW dataset belonging to one class, and a second set containing 390 non umpire images belonging to the other class. The Classifier 2 is designedfor the purpose of umpire pose classification or event detection. This classifier is trained on 390 umpire images from the SNOW dataset belonging to five

classes of events such as Six, No Ball, Out, Wide and No Action.

5.2. Phase 2 – Video Summarization We propose a system to summarize cricket videos by detecting important events that are signaled by the umpire. The saved classification models are combined into one system to realize this. The steps involved are as follows: pre-processing, feature extraction, umpire detection, event detection, frame accumulation and video summary generation. The input video is processed by extracting the frames sequentially. The frame rate of each video is 25 frames per second (fps). The following steps are performed sequentially for every frame in the video. Each frame is treated as a test image for the classifiers. Intensity normalization is performed as a preprocessing step. Then, the bottleneck features are extracted for these images using the pre-trained networks. These features are first tested on Classifier 1 to detect the presence of an umpire. If the frame is classified as an image belonging to the umpire class, then the features are carried forward, else the frame is discarded and the subsequent frame is processed. In the next step, the features are tested on Classifier 2 to detect the pose of the umpire and detect the event. If the frame was classified as an image belonging to one of the four classes of Six, No Ball, Out or Wide, then the processed frame is accumulated for generating the video summary. If the classified frame belongs to the no action class, then the frame is discarded as this is not relevant for event detection. Frames are accumulated into four individual sequences, each belonging to one of the categories of the detected pose of the umpire. Once all the frames are processed, the accumulated sequences are merged to generate the video summary

Processing each frame of a video in sequence is computationally expensive. Also, frame accumulation increases the memory footprint during runtime and is proportional to the length of the video. To overcome this memory overhead, a fixed size buffer is introduced in the frame accumulation and video summary algorithm. A buffer of size 250 frames is used to accumulate the incoming frames in sequence. For a frame rate of 25 fps this buffer can hold a video clip of 10 second duration. This form of extraction can be viewed as a moving window applied on the original input video. Each frame goes through the entire pipeline of umpire detection followed by umpire pose classification. In this manner, all 250 frames are processed. Based on the final classification result, these frames are accumulated into four individual frame sequences along with the detected class label for each frame. The total number of frames classified into each of the four classes is computed. A majority voter is used to decide the category of the summarized event.

For example, if the 250 frames in the buffer contain a scene of an event such as Out, then it is likely to contain frames of an umpire signaling Out. These frames, if correctly classified by Classifier 1 and Classifier 2, are expected to be accumulated more in number into the Out of category sequence than other categories. In the ideal case, other category sequences should not contain any frame for a scene containing Out. Based on a majority vote, these 250 frames are merged into a short video to summarize the Out event. The next 250 frames of the input video are processed in a similar manner in subsequent iterations. In a typical cricket video, it can be observed that the events such as Six, No Ball, Out and Wide last a duration of at least 10 seconds. Hence, a

moving window of 250 frames was empirically chosen to be ideal to detect and summarize the events.

6. EXPERIMENT AND ANALYSIS

6.1. CLASSIFIER DESIGN RESULTS

The classifiers are designed based on features extracted from the last fully connected layer of Inception V3 network, the first (fc1) and second (fc2) fully connected layers of the VGG19 network. Classifier 1 and Classifier 2 are trained on 80% of the dataset and the remaining 20% of the dataset is used for testing the classification performance. These classifiers are validated based on a 10- fold cross-validation and Jack-Knife or leave-one-out validation on the training data. The test accuracy is calculated based on the remaining 20% of the unseen data. The classification results are tabulated in Table 1. Classifier 1 is trained for umpire detection. From the results, it is apparent that Classifier 1 has a good performance accuracy for all three feature extraction methods. The performance of the classifier on features extracted using Inception V3 and fc2 layers of VGG19 are almost similar. The highest accuracy is achieved for the SVM trained on features extracted from fc1 layer of the VGG19 network. Classifier 2 is trained on the five class SNOW dataset for umpire pose detection. The highest performance is achieved for the SVM trained on features extracted from the fc1 layer of the VGG19 network. This classifier was tested on novel images that are not part of the training data. It can be concluded that Classifier 2 is successful in classifying any image into one of the five classes of poses. But it lacks the ability to detect the presence of an umpire in the image as it has not been designed to do so. Hence Classifier 1 is used in the overall system pipeline to add the ability of filtering the umpire images from non-umpire images. Similar performance is observed for classifiers trained on bottleneck features extracted from Inception V3 network. Classifiers 1 and 2 are combined such that both are trained using the same feature extraction method. For example, Classifier 1 trained on features extracted using fc1 layer of VGG19 is combined with the Classifier 2 trained using the same feature extraction method. The combined system is evaluated for its effectiveness in summarizing the cricket video

Classifier	Features	Accuracies		
		10-fold	Jack-Knife	Test
1	Inception V3	96.97%	97.76%	94.23%
	VGG19 – Fc1 layer	97.75%	97.59%	96.15%
	VGG19 – Fc2 layer	96.47%	96.79%	94.87%
2	Inception V3	77.71%	77.56%	85.90%
	VGG19 – Fc1 layer	82.43%	81.09%	83.33%
	VGG19 – Fc2 layer	78.14%	81.09%	78.21%

Table 7.1:-Training and testing Accuricies

8. RESULT

Fig 9.4-Result Example 2

Fig 9.3-Result Example 1

9 CONCLUSION

We proposed a system capable of summarizing cricket videos in the form of highlights based on detecting important events from the pose of the umpire. A new dataset, SNOW, containing umpire images for events such as Six, No Ball, Out and Wide was introduced in this work. This dataset has been made publicly available. The classification results indicated that features extracted from pre-trained networks such as VGG19 and Inception V3 provide a good performance baseline for this dataset. The combined system has been tested on cricket videos and is successful in detecting the majority of the events present in the video. The preliminary results obtained for the SNOW dataset suggest that this dataset is effective for the application of cricket highlights generation

10 REFRENCES

[1] https://ieeexplore.ieee.org/document/8628877

[2]https://towardsdatascience.com/support-vector-machine-introduction-to-machinelearningalgorithms-934a444fca47

[3]https://www.geeksforgeeks.org/deeppose-human-pose-estimation-via-deep-neural-networks/

[4] https://towardsdatascience.com/human-pose-estimation-simplified-6cfd88542ab3

[5]https://www.researchgate.net/publication/274149202_Predicting_Ball_Flight_in_Cricket_from_an_Umpire's_Perspective

[6] https://www.geeksforgeeks.org/image-classifier-using-cnn/