# Sports Video Classification - A Review

**Narayana Naik P[1], Prof. Nishil Kumar P.P[2]**

Fourth Sem. M. Tech in Signal Processing and Embedded Systems, ECE Department, GCE Kannur, Kerala[1]

Assistant Professor, ECE Department, GCE Kannur, Kerala[2]

**Abstract:** Video classification is a boundless topic. The Sports video classification is fundamental one and has considerable importance for archiving digital content in broadcasting companies. It is the process of classifying unknown sports video into the type of sport being played. This review paper analyses the different methods such as Neural Net, Texture code cue, Principle Component Analysis (PCA) and automatic thresholding, Mel Frequency Cepstral Coefficient (MFCC), Support Vector Machine (SVM), Hidden Markov model (HMM), CNN and RNN based Spatial and Temporal analysis in sports video classification based on sequential frame dataset. Convolutional Neural Network (CNN) is the basic one to extract features of images. Machine learning, as well as deep learning, approaches are popular for image recognition and classification. The simple and sophisticated method for video classification is "Transfer learning". It uses a simple and publicly available dataset. Transfer learning is the process of using a system or model, trained for one particular purpose to the other similar and specific purpose. VGG16 is a well-known pre-trained model for image classification. This model is considered for video classification.

**Keywords:** Neural Net, PCA, MFCC, SVM, HMM, CNN and RNN, Transfer learning.

## I. INTRODUCTION

In broadcasting media like Television, Streaming services, Internet, etc, videos are main domain and so many sports videos flow the data server day- to- day. In this review, sports videos are taken into account, because sports video is considered as simple process of video classification. Sports class is a limited discrepancy in all signs of its appearance. This approach is applied to design more accurate systems. Using video classification, one can avoid the boring manual searching for a particular type of video. Advanced feature extraction methods are used to the high-level information of video in system. The origin of video classification is lying in image classification applications. Many methods applied in video classification depends on the background of the image Deep Neural Network (DNN) based models are commonly applied at present. DNN is an effective method to solve the complex problems in signal processing and computer vision field. People consider a bunch of actions and the surrounding elements to recognise a sports item. A system be built to classify several sports classes, based on image frames in sequence. This system needs to verify the image information in both spatial as well as temporal domain. The CNN is capable to extract spatial feature, whereas RNN uses internal memory elements to process and store information throughout time. Combination of these two models are used to analyse sport video in different arrangement. In transfer learning, pretrained CNN models are used. The network has pre-trained on extensive data set, like ImageNet.

## II. LITERATURE REVIEW

A. Automatic Sports Classification using NeuralNet & Texture Code cue and Combination of these two.

Kieron Messer, William et al. [1] describes the method in automatically classifying the sports being played. The technique is the notion of "cues", is the semantic meaning to low-level features computed on the video. The system to be trained for the output ("cue") of the feature detected with a cram of people is identified in the scene of texture. These cues are incorporated with the contextual reasoning system to get high-level information; that is the class of sport played.

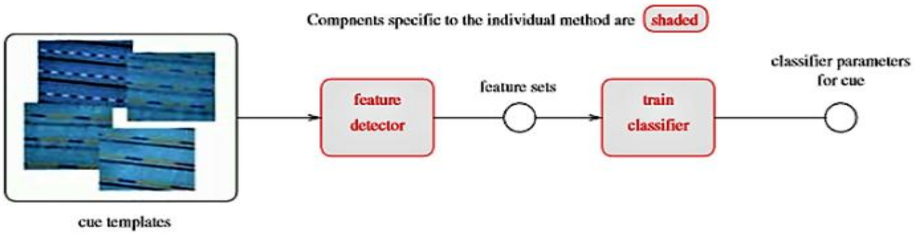
Compnents specific to the individual method are shaded
feature detector
feature sets
train classifier
classifier parameters for cue
cue templates

Fig. 1. Training of cue – detector [1]

The cue-detector has been trained as shown in Fig. 1. And the second stage training is needed to generates two pdf function $p(m|C)$ and $p(m|\bar{C})$, shown in Fig. 2. These pdf functions are the frequency of the output measurements, m, from a cue-detector

“C”, whether present or absent from the scene. For training as well as testing the system, frames of video were split into different sets viz, a cue-detector training set; the sports classifier training set and a test set.

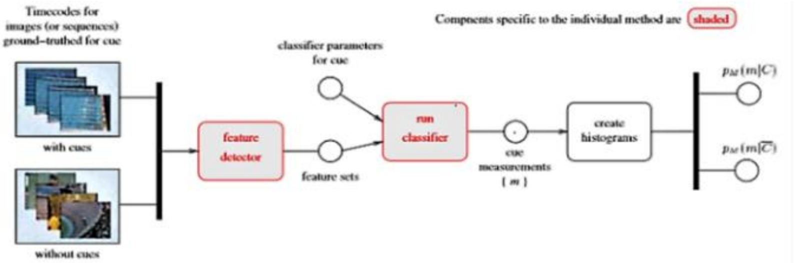

Fig. 2. pdf generation for Cue – detector [1].

Table (I), gives the details of ‘Cue’ methods used and ‘Cue’ detector.

Table I: CUE - Methods and CUE-Detectors Built

| Cue Method | Cue Detectors |
|---|---|
| Neural Net | Athletics Track, Boxing Ring, Indoor Cycling Track, Ocean Show Jumping Surface, Swimming Pool, Tennis Surface Grass, Blue Sky |
| Texture Cod | Boxing Ring, Cycle Track, Ocean Mid Shot Running Track, Distant Running Track, Medium Running Track, Near Swimming Lanes, Tennis Court |

B. Sports Type Classification using Signature Heat Maps

To control the privacy issues, Rikke Gade and et al. [2] applied thermal imaging. In this method, IR source produces images as show in figure 3, where pixel values represent the observed temperature. PCA (Principal component Analysis) is used to calculate the direction with considerable variance in samples by pooling all samples trained. The PCA will have non-zero Eigen values equal to the number of samples trained minus one. This work related to indoor games like, indoor soccer, basketball, volleyball, and badminton.

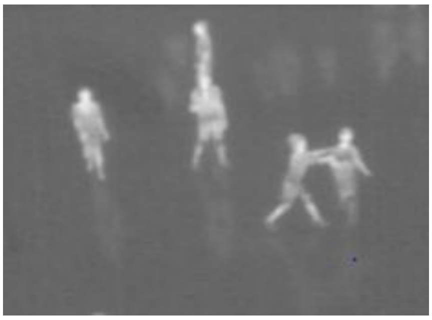

Fig. 3. IR source Imaging of Indoor Game [2]

Fischer’s Linear Discriminate (FLD) is preferred for classification. Fischer’s Linear Discriminate seeks the directions that are efficient for discrimination between the classes. The classification is performed based on the count of players in sports arenas. Automatic threshold matching method used to calculate the threshold that maximises the total entropy. It is illustrated in Fig. 4.

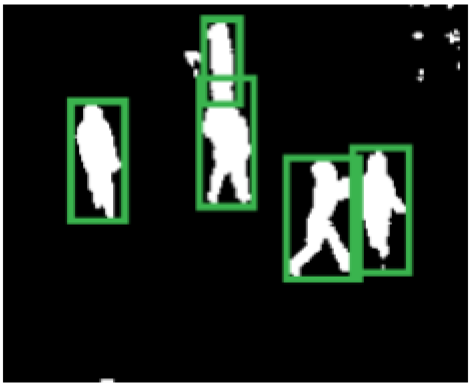

Fig. 4. Automatic Threshold method

C. Audio-visual classification of sports videos using the Mel frequency Cepstral Coefficient (MFCC)and Neural Net
Rikke Gade and et al. [3] used audio and video feature in combination for sports video classification. Audio feature extraction by MFCC by means of short-term power spectrum of the sound. This is obtained by," linear cosine transforms of a log power spectrum". This is the nonlinear scale of Mel frequency. Video feature extraction is performed by Neural Net. PCA will reduce the feature dimensions space to 10. The work performed in thermal videos of indoor soccer, basketball, volleyball, and badminton. A 'k-nearest neighbour' classifier is applied for classification.

D. Sports Video Classification in Continuous tv Broadcasts
Pavel Camprt et al. [4] is in Sports video classification in continuous TV broadcasts, CNN fc7(a layer before the classification layer is fc7), PCA200 (Principal Component Analysis) for feature extraction and SVM (support vector machine) for sports video classification is used. In this method," supervised classification" is used.

E. Event Recognition and Classification in Sports Video
Vijayan Ellappan et al. [5] in his "Event Recognition and Classification in Sports Video". Hidden Markov Model (HMM) is used in Part of-sound (POS) investigation and it is connected to games video titles and depictions. State transition machine is utilized to recognize the class. The figure (5) shows the block diagram representation for Event Recognition and Classification.

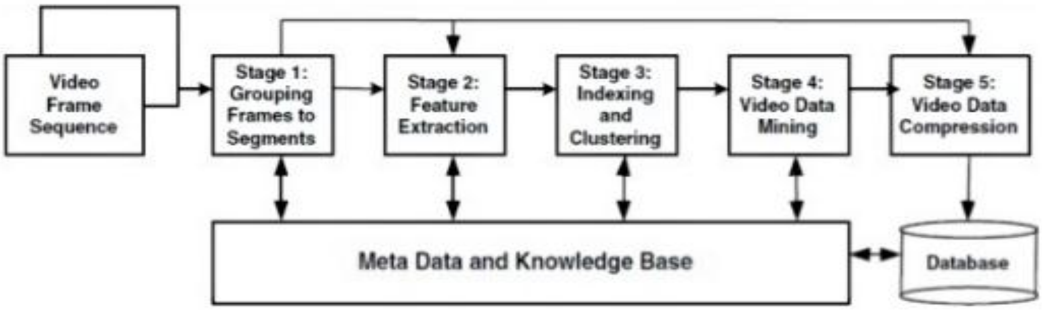
Video Frame Sequence
Stage 1: Grouping Frames to Segments
Stage 2: Feature Extraction
Stage 3: Indexing and Clustering
Stage 4: Video Data Mining
Stage 5: Video Data Compression
Meta Data and Knowledge Base
Database

Fig. 5. Block diagram of Meta data and knowledge-based feature extraction [5]

F. Sport Type Classification of Mobile Videos
Francesco Cricri, et al. [6] Main focus was to classify sports types for mobile videos. Authors considered three types of data such that captured video and audio as well as cameras and mic used in mobile phone to record the video and audio. Then developed multi class-SVM for their work.

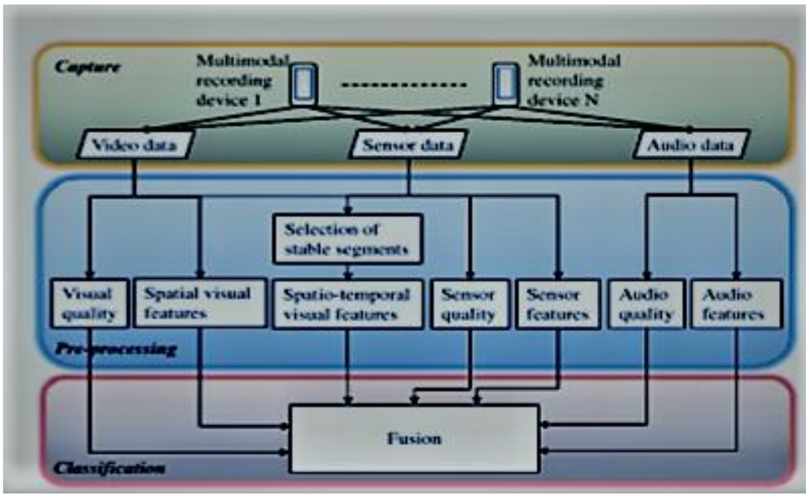

Fig. 6. Block diagram representation of video classification using multi class-SVM [6]

Spatial and temporal analysis is considered for sports type classification. Figure (6) shows the block diagram of representation Sports video Classification on Mobile device.

G. Sports Classification in Sequential Frames Using CNN and RNN
Mohammad Ashraf Russo, et al. [7]. Convolutional and recurrent network is used in combination for video classification. Sequential frames of datasets are considered for train the model.

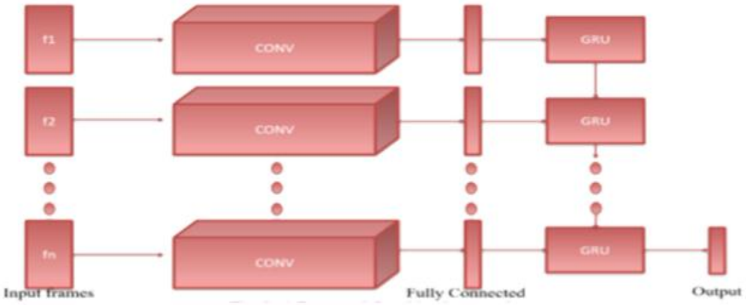

Fig. 7. Architecture of CNN with GRU [7]

CNN is used in spatial feature extraction of a signal. Whereas RNN is temporal feature extractor. The built-in memory elements with RNN is used for process temporal image information. GRU is upgraded class of standard RNN. This solves the problem of decaying features in an usual RNN. GRU are considered as update and Reset gate operations. The sequential RGB colour frames are input of network. Each frame is given input to each and separate convolutional layer. This method illustrated as in figure 7. The dataset used for this contains in number 300 video sequences derived from 50 different videos of recent sports events taken from YouTube. An example for sequential frame Dataset as in figure 8.

Fig. 8. Example of Sequential frame dataset

H. Classification of Sports Videos with Combination of Deep Learning Models and Transfer Learning [8]

Each CNN layer computing separate frame weights. A ReLU Unit to avoid the negative value if any in output of CNN layers (rectified linear unit) is used as activation function. The normalizing process is considered for standardizing the inputs to a layer for each batch. This stabilizes the learning process and reducing the number of training epochs required to train deep networks. A fully connected (FC) layer is used in between CNN and RNN unit. To prevent over fitting problem, a dropout of 0.5 was applied to get good results. For the recurrent part, the GRU has used as in figure 7.

Mohammad Ashraf Russo, et al. [8]. CNN & RNN and VGG16 is used in video classification. Transfer learning is the process of machine learning. It is defined as taking a model or solution trained on one task and re-using it to solve another similar task [8]. In transfer learning, CNN models used and pre-trained on ImageNet. VGG16 is a model of 16 layers of CNN with 3x3 convolution filter and for down sampling purpose max pooling of 2X2 with stride 2 is used. And 3 FC layers at the end of CNN for 1000-way feature extraction. SoftMax layer process is used for classification at the end of the network.

The VGG16 model is shown in figure 9. VGG16 is a Pre-trained model on the ImageNet dataset. The feature extractor that is suitable to classify images, can be used to video classification in transfer learning process. The Convolutional neural networks developed by Visual Geometry Group at University of Oxford for winning the ImageNet Challenge 2014 in localization and classification tasks are known as VGG Net. Commonly known Two classes of VGG Net are VGG16 and VGG19 Nets. VGG adopts the simplest structure. Only 3x3 convolution and stride 2, 2x2 max pooling are used throughout the whole network. VGG also shows the depth of the network which plays an important role in the process.

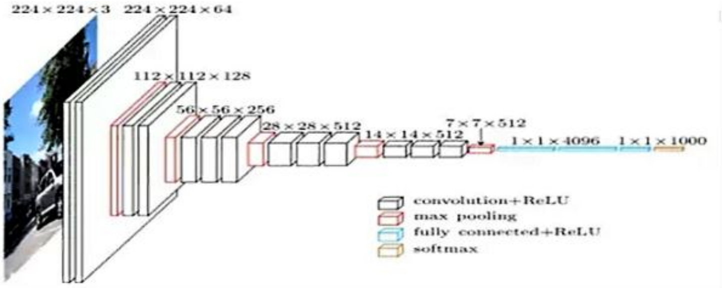

Fig. 9. VGG16 Architecture.

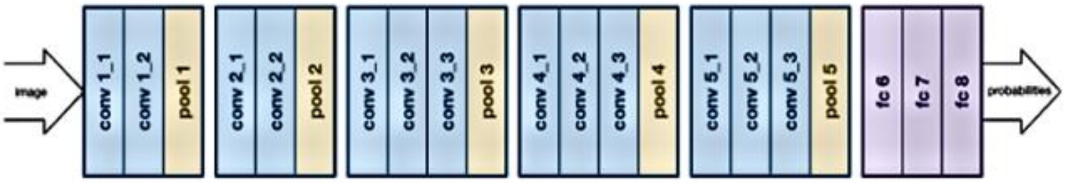

Fig. 10. Block representation of VGG16

Deeper networks provide better results. Usually VGG Net is big, is one of the draw backs of VGG net. It contains around 160M parameters. Most of the parameters are consumed in the Fully connected layers. VGG 16 Model having, 16 convolutional layers and it is very uniform architecture. Figure 9 shows the detailed view of VGG16 architecture. And figure 10 shows the detailed view of VGG16 block diagram representation. The VGG model is pre trained on ImageNet.

ImageNet is a Dataset of over fourteen million images belonging to 1000 classes. VGG architecture comprise a pre-processing layer. RGB colour image with pixel values in the range of 0-255 are taken as input. The value is calculated over the entire ImageNet Dataset. An example of image dataset is shown in figure 11. Down sampling is performed directly on convolutional layers which have a stride of 2. At the end of VGG network, a global average pooling layer and a 1,000-way fully-connected layer with SoftMax are used. SoftMax function gives the output vector that represents the probability distributions of a list of potential outcomes.

The data set used in image classification is also be used for video classification in transfer learning approach. In this process, sequential frames of videos not required. Hence storage of long videos are not required for training purpose. Such an example dataset is shown in figure 11.

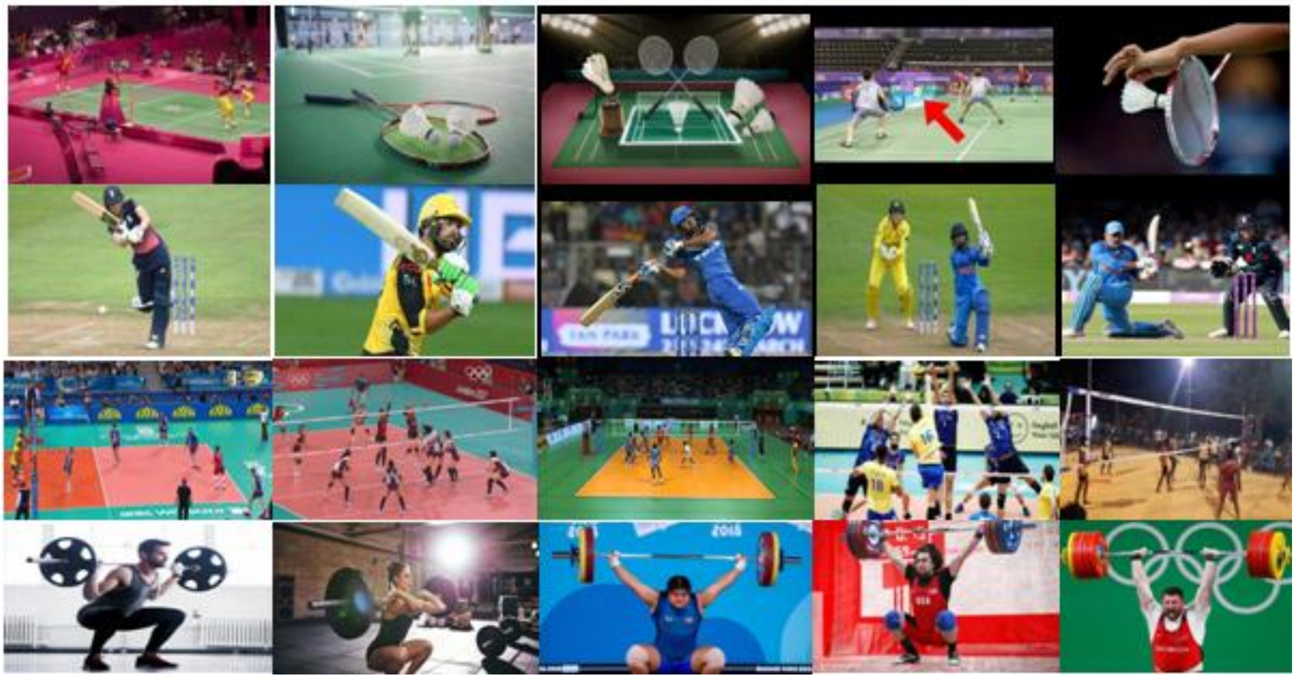

Fig. 11. Example image dataset for transfer learning.

## III. APPLICATIONS

Sports video classification is useful in searching for a particular sports video, analyse new tactics, videos on interest, etc. General video classification is broadly applicable in archiving of digital data, general video scene understanding, Videos on demand, Video ban, etc.

## IV. CONCLUSION

High-level features are required for the network to classify sports based on human actions and environmental scene context. CNN architecture has the ability of learning powerful features from weaker databases. CNN is having superior feature, is also applied in machine learning. As per the study of, R. Gade.et al. and V. Ellappan.et al., the video classification is based on both video and audio features. As per the study of, F. Cricri.et al., Video, audio and the sensor difference are also considered for classification of video. As per the study of M. A. Russo, et al., Video classification is performed by considering sequential frames of video using CNN and RNN. In their study by M. A. Russo, et al. the approach of transfer learning is explained. In this method, a pre-trained model VGG 16 for image classification is used. In all sense, the video classification requires more features, so a strong dataset is needed. It is painful to collect, annotate and store the sequential frame dataset. The video classification is based on image analysis applications. Hence new methods to be implemented for the classification of videos by means of using smaller datasets like image dataset. In the present scenario, Transfer learning is found to be the best method for video classification.

## REFERENCES

[1] K. Messer, W. Christmas, J. Kittler," Automatic sports classification", Proc. IEEE Int. Conf. Pattern Recognition, pp. 1005-1008, 2002.

[2] R. Gade, T. Moeslund," Sports type classification using signature heatmaps", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 999-1004, 2013.

[3] R. Gade, M. Abou-Zleikha, M. G. Christensen, T. B. Moeslund," Audio-visual classification of sports types", 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 768-773, Dec 2015.

[4] P. Campr, M. Herbig, J. Vanˇek and J. Psutka," Sports video classification in continuous TV broadcasts," 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, pp. 648-652, 2014.

[5] V. Ellappan and R. Rajasekaran," Event Recognition and Classification in Sports Video," 2017 2nd Inter. Confe. on Recent Trends and Challenges in Computational Models (ICRTCCM), Tindivanam, pp. 182-187, 2017.

[6] F. Cricri et al.," Sports Type Classification of Mobile Videos," in IEEE Transactions on Multimedia, vol. 16, no. 4, pp. 917-932, June 2014.

[7] M. A. Russo, A. Filonenko and K.H. Jo, "Sports Classification in Sequential Frames Using CNN and RNN", IEEE International Conference, 2018.

[8] M. A. Russo, Russo Laksono Kurnianggoro Kang-Hyun Jo, "Classification of sports videos with combination of deep learning models and transfer learning". International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019