The Journal of Supercomputing (2020) 76:7242–7267
https://doi.org/10.1007/s11227-020-03155-8

A decision tree framework for shot classification of field sports videos

Ali Javed[1] · Khalid Mahmood Malik[1] · Aun Irtaza[2] · Hafiz Malik[2]

Published online: 16 January 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Automated approaches to analyze sports video content have been heavily explored in the last few decades to develop more informative and effective solutions for replay detection, shot classification, key-events detection, and summarization. Shot transition detection and classification are commonly applied to perform temporal segmentation for video content analysis. Accurate shot classification is an indispensable requirement to precisely detect the key-events and generate more informative summaries of the sports videos. The current state-of-the-art have several limitations, i.e., use of inflexible game-specific rule-based approaches, high computational cost, dependency on editing effects, game structure, and camera variations, etc. In this paper, we propose an effective decision tree architecture for shot classification of field sports videos to address the aforementioned issues. For this purpose, we employ the combination of low-, mid-, and high-level features to develop an interpretable and computationally efficient decision tree framework for shot classification. Rule-based induction is applied to create various rules using the decision tree to classify the video shots into long, medium, close-up, and out-of-field shots. One of the significant contributions of the proposed work is to find the most reliable rules that are least unpredictable for shot classification. The proposed shot classification method is robust to variations in camera, illumination conditions, game structure, video length, sports genre, broadcasters, etc. Performance of our method is evaluated on YouTube dataset of three different genre of sports that is diverse in terms of length, quantity, broadcasters, camera variations, editing effects and illumination conditions. The proposed method provides superior shot classification performance and achieves an average improvement of 6.9% in precision and 9.1% in recall as compared to contemporary methods under above-mentioned limitations.

Keywords Decision tree · Rule-based induction · Shot classification · Shot transition detection · Sports videos

✉ Ali Javed
alijaved@oakland.edu

Extended author information available on the last page of the article

1 Introduction

Sports broadcasters generate enormous amount of video content due to the massive viewership of different sports around the globe. According to the statistics, more than half of the worlds population (around 3.6 billion) watched the 2018 mens soccer world cup [1], whereas the viewership of the 2019 mens cricket world cup reached to 2.6 billion worldwide [2]. We experience similar rise in viewership pattern in other sports around the globe. Manual analysis and processing of such enormous video content is challenging. Therefore, there is a strong need to develop automated methods for effective management of such massive sports videos available in the cyberspace. Sports videos content has been investigated for many years to provide various applications such as video content retrieval [3], indexing [4], browsing [5], shot classification [6–8] and summarization [9–11].

Shot classification is performed to segment the input video into different shots based on the camera views (i.e., long, medium, close-up, etc.). Effective shot classification is required to detect the key-events in sports videos that are then used to generate the highlights. Broadcasters are unable to analyze the sports videos for various applications in real-time due to semi-automated solutions. We aim to process and analyze the sports videos at real-time that will facilitate the coaches and players to review the recent performances during breaks/half-times in the game.

Video segmentation is a preliminary step in sports video analysis that is performed to detect the transitions between different video shots. Shot transition/boundary detection is performed to achieve temporal video segmentation that splits the video into separate shots and thus makes the process of video analytics more convenient. Shot boundary detection can be performed in various ways such as pixel-wise difference comparison, histogram difference comparison, edge change ratio comparison, etc. Pixel difference-based approaches [12, 13] are sensitive to camera motion that makes it less feasible for shot detection in sports videos where we experience extensive camera motion frequently. Performance of pixel difference-based techniques for shot boundary detection in sports videos are not up-to the mark and urge researchers to find more effective solutions. Edge change ratio-based approaches [14] are also used to detect shot boundaries by comparing edge change ratio between successive frames. The performance of edge-based approaches degrades in case of minor difference in edge change ratio between the two frames belonging to different shots. To address the limitations of pixel difference-based and edge change ratio-based approaches, comparative difference of histogram-based approaches [4, 15] has also been proposed. Histogram difference comparison between successive frames is a better approach for shot boundary detection in sports videos as it is independent of the camera motion.

Shot classification is usually performed after shot boundary detection to infer the semantics that are useful to classify the scenes and detect high-level events in the input videos. Existing shot classification approaches categorize the views into long, medium and close-up/out-field shots.

Existing approaches [15–18] on shot classification have used various features such as color, shape, edges, textures, motion, etc. Among visual features, color-based features are the most important for sports video contents analysis. For example, grass color pixel ratio is the most commonly used color feature for shot classification in sports video. Color pixel ratio was used in [15] to classify the shots among long, medium, and close-up views. Grass color pixels ratio can offer reasonable results to detect close-up shots as its low value indicates close-up shots, however, medium shots with high grass pixel ratio can be misinterpreted as long shots. This method [15] used grass color pixel ratio distribution of selected segments of frames to train a Bayesian classifier to distinguish between medium and long shots. In [17], grass color pixels were used to apply rule-based thresholding for long shots detection in soccer videos. This method [17] has limited utility due to its ability to determine only one category of shot. Grass color pixel ratio was also used in [18] to classify various shots for soccer videos.

Existing approaches [19–21] for shot classification have also used motion features in combination with color features for shot classification in the input sports videos. In [19], color and motion features were used to create an 11-dimensional feature vector, consisting of grass pixel color ratio and background motion features, that was fed to C4.5 decision tree for shot classification. Edge and motion features were used in [20] to train a support vector machine to classify various shots for tennis videos. A feature vector consisting of five edge distribution detectors and two optical flow features was created and fed to SVM for shot classification. These approaches have dependency on the motion variation of various shots. Similarly, color features were used in combination of texture features using radial basis decomposition and Gabor wavelets on SVM to classify the scene in field sports videos [21].

Shot classification problem has also been addressed using the low-level features in combination with mid-level features. In [11] a unified framework was proposed through employing supervised learning to perform a top down shot classification for sports videos. A nonparametric feature space analysis was performed to map low-level features consisting of color, texture, motion, and shot length into mid-level attributes such as motion entropy, active region, field shape, camera motion patterns, and shot pace that were then used for shot classification in sports videos. In [7], dominant color was used in combination of player size to classify the shots of soccer videos in long, medium, and close-up shots. Finally, Bayesian network was employed to detect the key-events for video summarization. Similarly, in [22] low-level features were used in combination of mid-level features to classify the shots in soccer videos. More specifically, in-field shots were filtered from the input video in the first stage. Later, a number of connected components and shirt color in vertical and horizontal strips were extracted from the in-field shots and used to train the SVM for shot classification.

Existing techniques also combined domain knowledge with low-level features to classify the shots in sports videos. In [23], play field color distribution and histogram were used in combination with domain knowledge to index key-frames for scene classification of tennis and baseball videos. Histogram difference comparison was used to detect the serving scenes in tennis videos. Face detection was used for close-up scene classification. Hue and saturation values were used to determine the skin color. If the detected skin color pixels of a region exceeded a specified threshold, then it was declared as a close-up shot. This method [23] has a limitation of false skin pixels detection for clay courts region in tennis videos and pitch field region in baseball videos. Likewise, Chen [24] used field color distribution, color histogram similarity, and color-based object location verification to detect scenes and events in tennis and baseball videos.

The ability of deep learning frameworks to automatically compute optimal features and train the network for classification has motivated research community to explore its benefits for shot classification. In [6], a deep learning feature fusion-based method was proposed to classify the shot types using the camera zoom and out-field information. CNN was used in [16] to classify the basketball videos into long and close-up shots.

Existing state-of-the-art [6, 16–19] has several limitations. The genre-specific methods [17–19, 25] have limited applicability, and dedicated algorithm implementation is required to support any other sports genre. Moreover, some approaches [15, 17] only use either few low- or high-level features to obtain reasonable accuracy. However, recent developments suggest that the combination of low- and high-level features achieves significant performance boost. Additionally, many existing approaches [6, 16] are computationally intensive and unsuitable for real-time applications. Existing shot classification methods also have dependency on camera variations, illumination conditions, game structure, shot speed, sports genre, broadcasters, etc. To address all these aforementioned challenges, we propose a decision tree framework for shot classification of field sports videos.

The major contributions of the proposed method are as follows:

- Effective fusion of different low-, mid-, and high-level features to design a decision tree framework for accurate shot classification of field sports videos.
- Rule-based induction is applied to create various rules from the decision tree to classify the video shots into long, medium, close-up, and out-of-field shots.
- One of the contributions of the proposed work is to find the most reliable rules that are least unpredictable for shot classification of the sports videos.
- The proposed method is robust to illumination conditions, different ground and pitch fields, variations in camera and motion, shot speed, game genre and structure, etc.

The rest of the paper is organized as follows: Sect. 2 presents the proposed shot boundary detection method. Section 3 discusses the proposed decision tree framework for shot classification. Section 4 provides a comprehensive analysis of the results and different experiments that are designed to evaluate the performance of the proposed method. Finally, Sect. 5 concludes the proposed work and provides a brief discussion on the future prospects.

2 Shot boundary detection

Shot boundary detection is commonly applied to perform video segmentation by observing either abrupt or gradual transitions between consecutive frames. In our method, abrupt transitions are detected to partition the input sports videos into various shots. The general theme is to detect the difference between the frames represented by either the local features (e.g., SIFT, LBP, and kernel density estimation) [26, 27] or global features (i.e., color histograms, color correlograms and texture patterns) [28–30]. As the techniques employing local features for frame representation are sensitive to zooming and camera position, therefore, their practical applications are very limited. Hence, in the proposed work shot boundaries are detected by histogram comparison of luminance (grayscale) component between the consecutive frames. The process flow of the proposed shot boundary detection method is presented in Fig. 1.

The input sports video is transformed into grayscale representation during pre-processing. The histogram of each grayscale frame is computed. Frames signify different shots if their histograms illustrate a prominent difference over the grayscale intensities. Difference of the histograms is computed as (Eq. 1):

$$D_i = \sum_{i=1}^{B} |H_j(i) - H_{j+1}(i)| \tag{1}$$

where D_i represents the obtained difference while $H_j(i)$ and $H_{j+1}(i)$ represents the consecutive histograms with B bins. The value of D_i will be small (i.e., less than a threshold) for the frames having similar shots and vice versa. The intensity histogram difference shows minimal value in case of continuous video frames, illumination variations and prominent movements as compared to abrupt transitions where its values shoot. This feature makes the intensity histogram difference a perfect technique to detect cuts in videos by utilizing a suitable threshold value. In the proposed work, first- and second-order statistics are applied to calculate the threshold value as described in Eq. (2):

$$T_{\text{sbd}} = \mu_{\text{h}} + p_{\text{s}}\sigma_{\text{h}} \tag{2}$$

where μ_{h} and σ_{h} represents the first- and second-order statistics of the histogram of each frame respectively. The parameter is a positive integer value in the range of 3

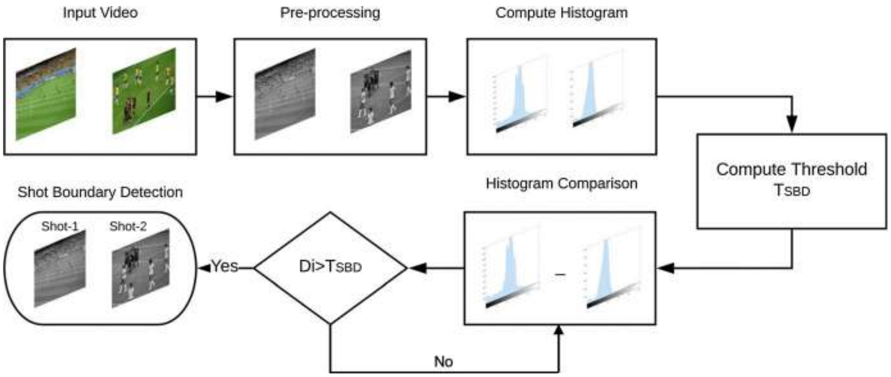


Fig. 1 Process flow of proposed shot boundary detection

to 6 in the implementation of the proposed work. The designed experiments have obtained the most optimal results in this range.

3 Shot classification

Shots classification is usually applied after shot boundary detection to label each shot. Shot class information is commonly used in combination with some low-level features to express useful semantics regarding the video content. Shots are usually categorized into long, medium and close-up views in general [15]; however, in case of sports videos, the shots can be categorized into various classes [31]. The proposed method classifies the sports videos into long, medium, close-up, and out-of-field/crowd shots as shown in Table 1. The detail of the proposed framework is presented here.

3.1 Grass and pitch field detection

The input sports videos are transformed from RGB to HSV color space [13] to take the advantage of processing the chrome and luminance components separately. Additionally, the distribution of color in HSV color space is uniform that makes the color processing more convenient in HSV color space. The Hue component is extracted from HSV color space for each frame and analyzed further to determine the grass field and pitch field pixels color. In the proposed method, a histogram is plotted for the hue component of each frame in the dataset videos during training. The dataset videos created for shot classification is diverse in illumination conditions (i.e., daylight, artificial light matches), length, sports genre, etc. It has been observed after analyzing the histograms of cricket and baseball videos that it contain two dominant ranges of peaks, one each for grass field and pitch field pixels. Hence, histograms of baseball and cricket videos are bimodal, whereas the histogram of soccer is unimodal as shown in Fig. 2; grass field pixel distribution is encircled with red dotted arrow, whereas pixel field pixel distribution is encircled with green arrow.

The histogram of the hue component is analyzed to determine the color distribution for grass and pitch field pixels during the training stage. We performed the histogram analysis of the entire video dataset to determine the search space of color bins for both the grass field pixels and pitch field pixels. We used 256 bins to generate the histogram of each frame of given video and done it for all videos

Table 1 Shots categorization

Shot category	Details
Long shot	Presents global view of the field
Medium shot	Presents zoomed view (i.e., full player body)
Close-up shot	Presents zoomed-in view (i.e., above-waist view of players)
Out-of-field/crowd shot	Presents audience/out-field view

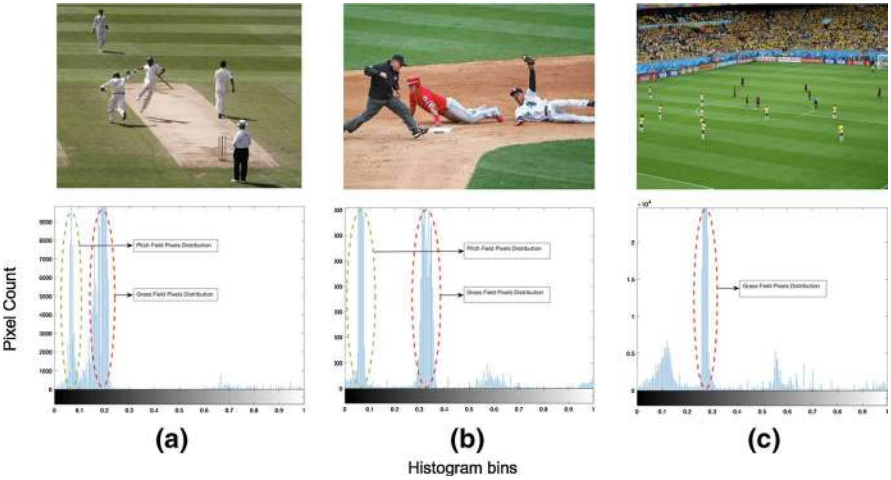


Fig. 2 Histogram peak analysis for ground field pixels and pitch field pixels of cricket (**a**), baseball (**b**), and soccer (**c**) videos

in the dataset. As mentioned earlier, that we obtain a unimodal histogram for soccer and bimodal for cricket and baseball due to massive number of pitch field pixels along with the grass field pixels. We plotted the average of these two bimodal histogram distributions for the entire video dataset. The ranges of bins 40–70 for grass field pixels and 10–35 for pitch field pixels are obtained through this initial histogram analysis conducted on all videos of our dataset. The proposed approach calculates the histogram peak index, g_{peak} and p_{peak} for grass and pitch field by analyzing the histogram of hue component in these observed ranges. These peaks are then used to compute the grass field and pitch field pixels distributions. The peaks are identified after computing the average of histogram peaks on all frames of different videos and sports genre of the dataset. The peak is included for averaging if the histogram peak exceeds the threshold T_{peak}. g_{peak} is detected at bin, b = 56, and p_{peak} is detected at bin, b = 26. Since the search space is small for T_{peak}, therefore, we performed a brute force technique to compute the threshold T_{peak}. We performed this experiment on different values of T_{peak} (i.e., 3000, 4000, 5000, 6000, 7000, 8000). We obtained the best results on T_{peak} set to 6000. Therefore, in the proposed work, we used $T_{\text{peak}} = 6000$ for experimentation.

In the proposed method, an interval ($g_{\min} \leq g_{\text{peak}} \leq g_{\max}$) is computed for grass field pixels, and second interval is defined for pitch field pixels that is ($p_{\min} \leq p_{\text{peak}} \leq p_{\max}$). $g_{\min}$ and $p_{\min}$ are identified by finding the lowest bins in the range of 40 to 70 and 10 to 35, respectively, that exceeds the threshold T_{peak} over all frames of various sports videos during training. Similarly, $g_{\max}$ and $p_{\max}$ are detected by finding the highest bin in the same ranges and threshold described above. The interval ($g_{\min} \leq g_{\text{peak}} \leq g_{\max}$) is detected at bins, $b = 48$ to $b = 64$ and ($p_{\min} \leq p_{\text{peak}} \leq p_{\max}$) at histogram bins, $b = 18$ to $b = 34$ in the proposed method.

3.2 Feature extraction for shot classification

Effective feature extraction is a key requirement for optimal shot classification. The proposed method uses some innovative features like upper body detector in combination of commonly used features, i.e., grass field color pixels, pitch field pixel ratio, edge pixels ratio, people detector, etc., to perform shot classification of sports videos.

3.2.1 Grass field pixel ratio (GFPR)

The dominant field color is a key attribute for video content analysis of field sports. The field color is dependent on the illumination conditions, different stadiums, and grass color. As described in Sect. 3.1 earlier, grass field pixels are computed by identifying an interval ($g_{\mathrm{min}} \leq g_{\mathrm{peak}} \leq g_{\mathrm{max}}$). Each frame of the input sports video is transformed in HSV color space and hue component is analyzed to compute the grass field pixels ratio (GFPR). We created a rule to determine the grass field counter g_{count}. More specifically, the hue value of each pixel is scanned and g_{count} is incremented if the corresponding pixel lie in the interval $[g_{\mathrm{min}}\ g_{\mathrm{max}}]$ as:

$$R_{\mathrm{GFC}} = \begin{cases} g_{\mathrm{count}}++, & \text{if } g_{\mathrm{min}} \leq I^{(i)}(r,c,1) \leq g_{\mathrm{max}} \\ g_{\mathrm{count}}, & \text{Otherwise} \end{cases}$$

where $I^{(i)}(r,c,1)$ represents the hue component of ith frame of the input sports video. Finally, GFPR is computed as follows:

$$\mathrm{GFPR} = \frac{g_{\mathrm{count}}}{p_{\mathrm{total}}} \tag{3}$$

where p_{total} is the total number of pixels in any given video frame.

3.2.2 Pitch field pixel ratio (PFPR)

The proposed method uses pitch field pixels ratio (PFPR) in combination with other features to classify between the long and medium shots in cricket and baseball videos. Since the grounds in cricket and baseball contains pitch area for batting and bowling, therefore, pitch field pixel color is used in the proposed work to classify various shots.

The pitch field color is a useful descriptor to classify various shots in cricket and baseball videos. Like ground field color, pitch field color also depends on different stadiums, and illumination conditions that change under different weather conditions.

As described in Sect. 3.1 earlier, pitch field pixels are computed by identifying an interval ($p_{\mathrm{m\,i\,n}} \leq p_{\mathrm{peak}} \leq p_{\mathrm{m\,a\,x}}$). Each frame of the input sports video is transformed in HSV color space and hue component is analyzed to compute the PFPR like GFPR. Again we create a rule to determine the pitch field counter p_{count}.

More specifically, each pixel is scanned to analyze the hue value, and p_{count} is incremented if the corresponding pixel lie in the interval $[p_{\min}\ p_{\max}]$ as follows:

$$R_{\text{PFC}} = \begin{cases} p_{\text{count}} + +, & \text{if } p_{\min} \le I^{(i)}(r, c, 1) \le p_{\max} \\ p_{\text{count}}, & \text{Otherwise} \end{cases}$$

Finally, PFPR is computed as follows:

$$\text{PFPR} = \frac{p_{\text{count}}}{p_{\text{total}}} \tag{4}$$

3.2.3 Edge pixel ratio (EPR)

Ground field pixels ratio solely is not sufficient to clearly distinguish between the close-up and out-of-field/crowd shots. False alarm rate increases significantly for close-up shots if GFPR is analyzed only to classify between the close-up and out-of-field shots. It has been observed that the out-of-field/crowd shots contain more edges as compared to close-up shots due to the presence of many people and objects. Therefore, edge pixel ratio is used in combination with GFPR in the proposed method to classify between the close-up and out-of-field/crowd shots. This fusion of features increases the accuracy of the proposed method in terms of accurate classification between the close-up and out-of-field/crowd shots. Canny edge detector [32] is employed to compute the edges in each frame. EPR is computed by comparing edge pixels to the total number of pixels in each frame.

3.2.4 Histogram of oriented gradients

Sports videos contain prominent footages of players and referees in medium and close-up shots/views. Therefore, people detector/detection (PD) is used in the proposed method to detect the full body view of the player/referee in the long and medium shots. Histograms of oriented gradient features [33, 34] are used to train the SVM classifier [35] that is then used to detect the players in an upright position in the input sports video.

Each video frame is initially processed by normalizing the gamma and color values during the pre-processing stage. Gaussian smoothing is applied on this normalized image followed by a 2-D derivative operator (2×2 Robert Cross gradient operator, and 3×3 Sobel operator) [36] to compute the gradient. Gradients are computed for each color channel and the one with the largest norm is taken as the pixels gradient vector. The orientation bins are created by making each pixel to compute a weighted vote for an edge orientation histogram channel. The votes are gathered into orientation bins over local spatial regions, labeled as cells. These cells are grouped into spatial blocks followed by applying the normalization on each block separately. Histogram of gradient feature vectors are extracted from the detector window tiled with a grid of overlapping blocks. The size of the detection window stride is selected as 8×8 in the proposed work. Finally, human classification is achieved by feeding the combined vectors into a linear SVM classifier.

3.2.5 Haar-based features

Close-up shots depict the above-waist view of players in the sports videos that present the facial exposure of the players prominently. Face detector (FD) is an important feature that can be used in combination with other features to detect the close-up shots. In the proposed method, a new descriptor, upper body detector (UBD) is also used in combination of face detector to effectively classify among various shot categories. Haar-based features [37] are used for face and upper body detection. Haar-based features consider adjacent rectangular areas at an explicit position in a detection window followed by adding the intensities in each area and compute the difference between these added intensities. This difference is then used to classify subsections of an image. It is a common observation that the eyes region in the face is darker than the cheeks region. Therefore, for face detection, Haar features consist of a set of two adjacent rectangles that lie above the eye and cheek region. The location of these adjacent rectangles is defined relative to a detection window that acts as a bounding box to the target object. These Haar-based features [37] are used to train a classifier. Integral image representation is used to compute features quickly. AdaBoost algorithm is applied to create a classifier by picking a small collection of features. The computation speed of the detector is increased significantly by integrating complex classifiers consecutively in a cascade structure.

3.2.6 Object scale (OS)

The scale of objects can be used effectively to analyze the sports videos for shot classification. Long view consists of players footage in small scale as compared to medium and close-up shots where the object scale appears larger. The proposed method transforms each input video frame into binary image. Connected component labeling via 8-connectivity is applied to label all the components. The connected components (blobs) of significantly huge size are filtered and used in the proposed method for shot classification. Each blob is compared against a specified threshold T_{blob}. If the size of a blob in terms of number of pixels exceeds T_{blob} then the corresponding blob is retained in the image else it is discarded. T_{blob} is set to 15,000 after extensive experimentation in the proposed method as optimal results are achieved on this threshold in our implementation.

3.3 Decision tree framework for shot classification

In the proposed work, a multi-layer decision tree is created using the feature set consisting of GFPR, EPR, PFPR, OS, FD, UBD and PD for shot classification. The reason to adopt decision trees for shot classification is twofold. First, the ability of decision trees to mimic the human thinking makes it easy to generate rules for data analysis and interpretation. Second, data preparation for decision trees is relatively easy and can handle both numerical and categorical data.

In addition, decision tree framework is computationally efficient as compared to other classifiers like SVM or CNN, which makes them more suitable for real-time applications. The structure of the decision tree is shown in Fig. 3.

Rule-based induction is applied to design various knowledge-based rules to classify the shots among long, medium, close-up and out-of-field shots. Entropy is computed against each rule to filter the rules that are least unpredictable in terms of classifying among various shot categories as described earlier. The input sports video is already partitioned into various shots via shot boundary detection as discussed in Sect. 2. Frames from each video shot are processed further to classify each frame among long, medium, close-up and out-of-field shots/views. Finally, a majority voting scheme is applied to label each video shot with the class of shot belonging to majority of frames. As it is evident from Fig. 3 that there exist many paths from the root node to the leaf nodes. Each path creates a rule that is analyzed to determine the predictability of each rule in terms of shot classification. The most reliable rules identified after entropy computation for shot classification are presented in this section.

The root node at level-0 (L0) analyzes the grass field pixels ratio. There exist several different possible outflows from the root node on the basis of GFPR. If the GFPR exceeds the threshold T_{LS}, then further features are analyzed at level-1 (L1) to determine the possibility of either long, medium or close-up shot. Upper body detector (UBD) is examined at L1 in case of GFPR exceeds T_{LS}. If UBD is true (i.e., upper body of person(s) detected), then the frame is classified as close-up view. In case UBD is false then face detector (FD) is analyzed at level-2 (L2) of the decision tree. If FD is true (i.e., face of person(s) detected) then the frame is also classified as a close-up view. In case FD is false then people/player detector (PD) is examined at level-3 (L3) of the decision tree. If PD lies below the threshold T_P then the frame is marked as a long view. In case PD exceeds T_P then object scale (OS) is analyzed at level-4 (L4) of the decision tree. If OS is true (i.e., object(s) of significant size detected) then the frame is classified as a medium

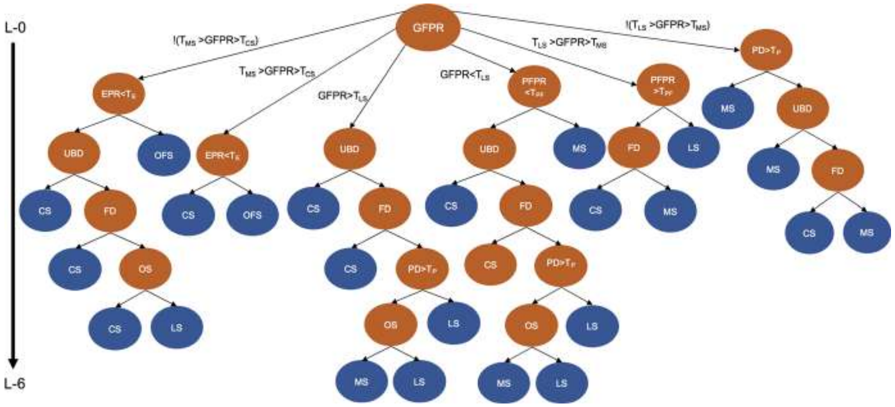


Fig. 3 Decision tree architecture for shot classification

view. In case OS is false (i.e., no object(s) of significant size detected) then it is labeled as a long view at level-5 (L5) of the decision tree.

For cricket and baseball videos, if PFPR lies below the threshold T_{PF}, then the frame is classified as a medium view frame else UBD is analyzed at L2. If UBD is true, then the frame is classified as a close-up view else FD is examined at L3. If FD is true, then it is also marked as a close-up view frame else PD is analyzed at L4 of the decision tree. If PD lies below the threshold T_P then the current frame is marked as a long view frame else OS is analyzed at L5. If OS is true, then the frame is marked as a medium view else as a long view frame.

In addition, if GFPR exceeds the threshold T_{MS} and lies below T_{LS} (i.e., $T_{\text{LS}} > \text{GFPR} > T_{\text{MS}}$), then PFPR is analyzed for cricket and baseball at L1. This is due to the fact that only cricket and baseball contain the pitch fields in the playing area unlike soccer that consists of grass field only. If PFPR exceeds the threshold T_P, then the frame is detected as a long view at L2 of the decision tree. In case if PFPR lies below T_P, then FD is examined at L2 of the decision tree. If FD is true, then the frame is marked as a close-up view else as a medium view frame.

For all of the three sports genre including soccer, if GFPR meets this condition ($T_{\text{LS}} > \text{GFPR} > T_{\text{MS}}$), then PD is analyzed at L1. If PD exceeds threshold T_P, then the frame is classified as a medium view frame at L2. In case PD lies below T_P then UBD is examined at L2 of the decision tree. If UBD is true, then it is marked as a medium view frame else FD is analyzed at the next level. If FD is true, then the frame is labeled as a close-up view else as a medium view frame at L4 of the decision tree.

Moreover, if GFPR exceeds the threshold T_{CS} and lies below T_{MS} (i.e., $T_{\text{MS}} > \text{GFPR} > T_{\text{CS}}$), then edge pixel ratio (EPR) is analyzed to discriminate between the close-up and out-of-field views. If EPR lies below the threshold T_E, then the frame is classified as a close-up view else as an out-of-field/crowd view. This is due to the fact that out-of-field/crowd views contain maximum distribution of edge pixels. Additionally, if GFPR do not lie in the range of T_{CS} and T_{MS} i.e., $!(T_{\text{MS}} > \text{GFPR} > T_{\text{CS}})$, then again EPR is investigated to discriminate among different views. At L2, if EPR exceeds T_E then out-of-field view is marked for the current frame else UBD is examined at the next level. If any of the UBD, FD, or OS is true at L2, L3, and L4, respectively, then the frame is labeled as a close-up view. In case all of UBD, FD, and OS are false then the current frame is detected as a long view frame as shown in Fig. 3.

3.3.1 Long view/shot classification

Long view/shot is classified using some low-level features (GFPR and PFPR) in combination with mid-level features (OS) and high-level features (FD, PD, and UBD). GFPR and PFPR are calculated by analyzing the color values of the hue component of ground field and pitch field pixels as described in Sect. 3.1. Rule-based induction is applied to generate five rules for long shot classification as shown in Fig. 3. Entropy is computed to determine the most predictable rule that can classify the long shot with better accuracy. The least unpredictable rule selected for long shot classification is as follows:

$$R_{\mathrm{LS}} = \begin{cases} \mathrm{GFPR} > T_{\mathrm{LS}} \wedge \mathrm{PD} < T_P \wedge \mathrm{FD} = \mathrm{False} \wedge \mathrm{UBD} = \mathrm{False} \wedge \mathrm{OS} = \mathrm{False} \\ \vee \\ \mathrm{GFPR} < T_{\mathrm{LS}} \wedge \mathrm{PFPR} > T_{\mathrm{PF}} \wedge \mathrm{PD} < T_P \wedge \mathrm{FD} = \mathrm{False} \wedge \mathrm{UBD} = \mathrm{False} \wedge \mathrm{OS} = \mathrm{False} \end{cases}$$

If GFPR exceeds T_{LS}, PD lies below T_P, and UBD, FD, and OS are false, then the corresponding frame is labeled as a long view frame. These conditions are applicable to all three sports genre. In addition, if PFPR exceeds T_P, PD lies below T_P, and UBD, FD, and OS are false, then the corresponding frame is also labeled as a long view frame. These conditions are applicable to cricket and baseball videos as soccer field do not contains the pitch field area.

For long shot classification, GPFR and PFPR are analyzed against the thresholds T_{LS} and T_{PF}, respectively. GFPR must exceed T_{LS} and PFPR must exceed T_{PF} as described earlier. T_{LS} is set to 0.4, and T_{PF} is set to 0.09 in the implementation of the proposed work. These values are adopted due to the optimal results obtained on these values after detailed experimentation. Moreover, it has been observed after watching massive amount of different field sports videos that long shot are usually comprised of majority of grass field color distribution.

Shown in Fig. 4 are the input sports video frame along with the grass field pixel distribution and pitch field pixel distribution of the long shots of cricket, baseball and soccer videos. GFPR and PFPR distribution are represented in the foreground

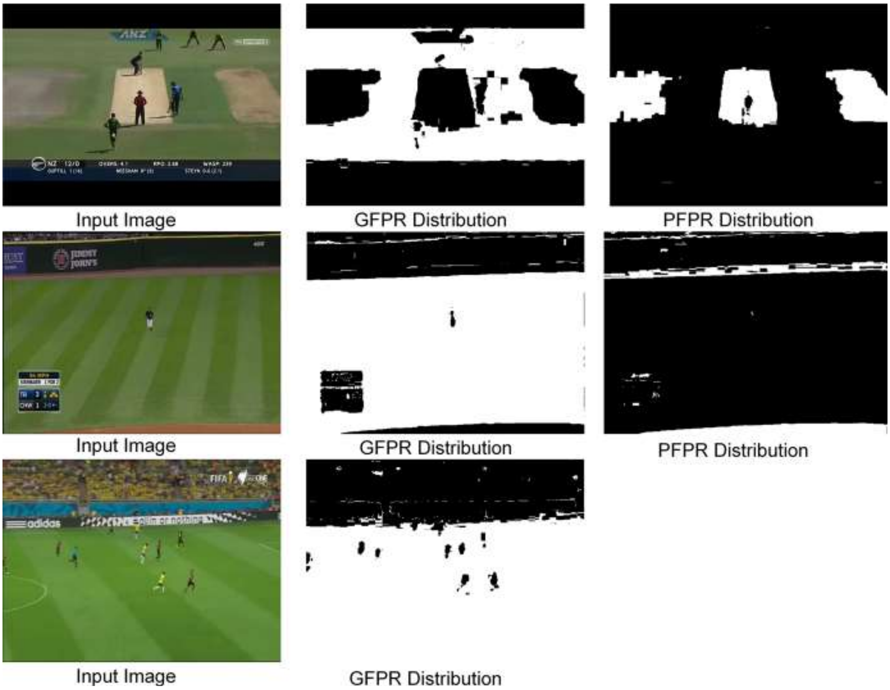


Fig. 4 Row-1: long shots for cricket, row-2: long shots for baseball, row-3: long shots for soccer

(i.e., white color in the frame) and rest of the content in the background (i.e., black color in the frame). We can clearly observe from Fig. 4 that the proposed method effectively computes the regions of grass field and pitch field pixels in the sports video. Since the soccer field do not contains the pitch field area, therefore, PFPR is not computed for soccer videos.

3.3.2 Medium view/shot classification

Medium view/shot is classified using the features of GFPR, PFPR, PD, FD, and UBD. Rule-based induction is applied in the same way to generate seven rules for medium shot classification as shown in Fig. 3. Entropy is computed to determine the most predictable rule that can classify the medium view with better accuracy, and achieves superior detection performance among the others. More specifically, the entropy of these seven rules is computed to filter the rule with lowest entropy value and later used for medium view classification.

The proposed method assigns a label of medium view to the current frame if the GFPR satisfy this condition (T_{MS} < GFPR < T_{LS}), PFPR is lower than T_{PF} and FD is false. These conditions are applicable to both cricket and baseball. In addition, the current frame is also labeled as medium view if GFPR lies between T_{LS} and T_{MS}, PD is greater than TP, FD is false, and UBD is true. These conditions are applicable to all three genre of sports videos. For medium shot classification, the selected rule is as follows:

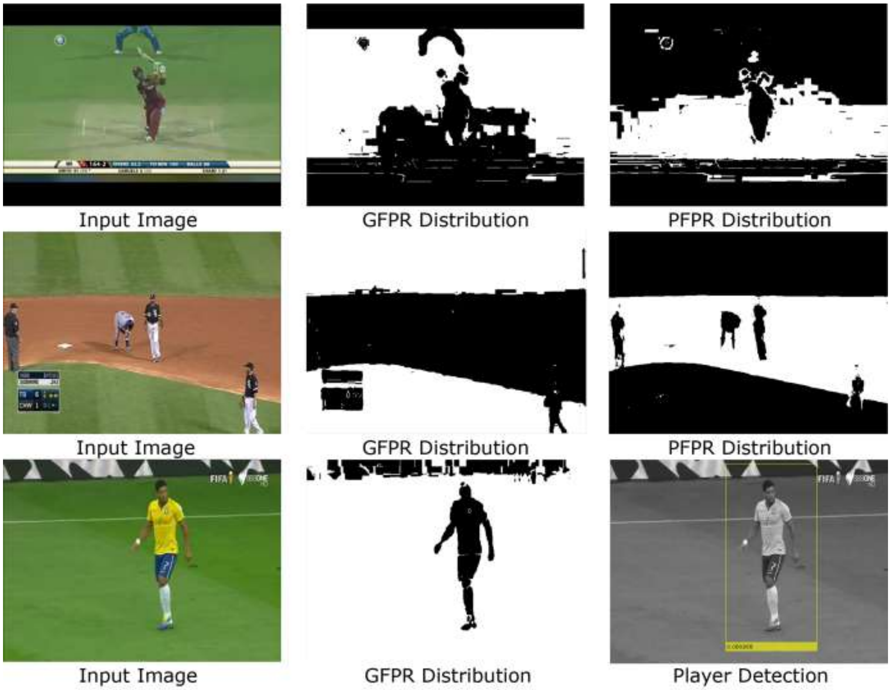
$$R_{MS} = \begin{cases} \mathrm{T_{MS}} < \mathrm{GFPR} < T_{\mathrm{LS}} \wedge \mathrm{PFPR} < T_{\mathrm{PF}} \wedge \mathrm{FD} = \mathrm{False} \\ \vee \\ \mathrm{T_{MS}} < \mathrm{GFPR} < T_{\mathrm{LS}} \wedge \mathrm{FD} = \mathrm{False} \vee \mathrm{PD} > T_P \vee \mathrm{UBD} = \mathrm{True} \end{cases}$$

For medium shot classification, GFPR is analyzed against the thresholds T_{LS} and T_{MS}. T_{MS} is set to 0.25 in the implementation of the proposed work as the distribution of grass field in medium shots is usually lower as compared to long shots. Moreover, the proposed method has achieved better results on this parameter value for T_{MS}.

Figure 5 shows the input sports video frame along-with the grass field pixel distribution and pitch field pixel distribution frames of the medium shots for cricket and baseball. The GFPR and PFPR distribution in the video frames are represented in the same way as described earlier (Sect. 3.3.1). For soccer videos, grass field pixel distribution and player detector snapshots are presented in Fig. 5.

3.3.3 Close-up view/shot classification

For close-up shot classification, EPR is used in combination of GFPR, FD, UBD, and OS. Ten rules are created from the decision tree by applying the rule-based induction to classify the close-up shots as shown in Fig. 3. The entropy of these ten rules are computed to select the rule with lowest entropy value and further used for close-up shot classification. The selected rule is listed below.



Input Image GFPR Distribution PFPR Distribution

Input Image GFPR Distribution PFPR Distribution

Input Image GFPR Distribution Player Detection

Fig. 5 Row-1: medium shots for cricket, row-2: medium shots for baseball, row-3: medium shots for soccer

$$R_{\mathrm{CS}} = \begin{cases} \mathrm{T}_{\mathrm{CS}} < \mathrm{GFPR} < T_{\mathrm{MS}} \wedge \mathrm{EPR} < T_E \\ \qquad\qquad\qquad\qquad \vee \\ \mathrm{EPR} < T_E \wedge \mathrm{FD} = \mathrm{True} \vee \mathrm{UBD} = \mathrm{True} \vee \mathrm{CC} = \mathrm{True} \end{cases}$$

Each frame is examined and classified as a close-up view if the GFPR satisfy this criteria (T_{CS} < GFPR < T_{MS}) and EPR is lower than the threshold T_E. Additionally, if EPR lies below T_E and FD, UBD, and OS are true, then the corresponding frame is also labeled as close-up view.

For close-up shot classification, GFPR is analyzed against the thresholds T_{CS} and T_{MS}, whereas, EPR is compared against the threshold T_E. T_{CS} is set to 0.1 and T_E is set to 0.07. The reason to use lower values for EPR and GFPR is because of their distribution in close-up shots usually exist in small quantity in the sports videos. Moreover, optimal results are achieved on these parameter values after the detailed experimentation.

Figure 6 shows the input video frame, grass field pixel distribution, edge pixel distribution, face detector, and upper body detector of the close-up shots of cricket, baseball, and soccer. As it can be observed from Fig. 6 that the proposed method effectively calculates the GFPR, EPR, face and upper body detection in the close-up shots of cricket, baseball, and soccer videos.

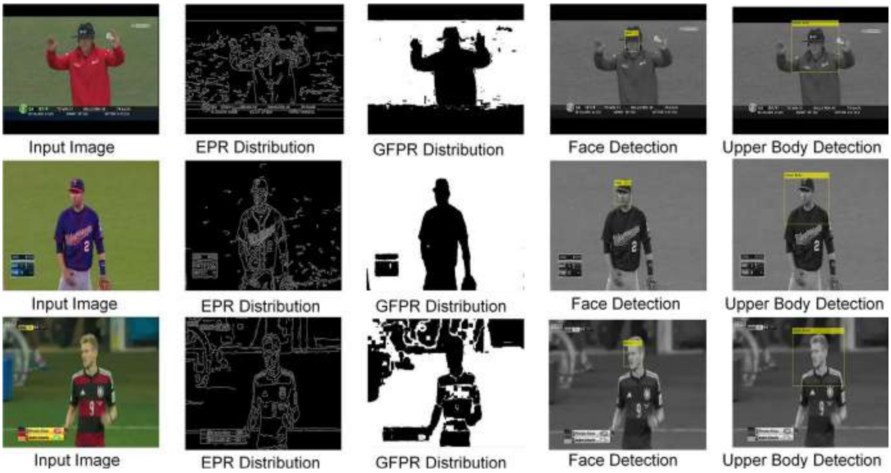


Fig. 6 Row-1: close-up shots for cricket, row-2: close-up shots for baseball, row-3: close-up shots for soccer

3.3.4 Out-of-field view/shot classification

The proposed method includes the out-of-field and crowd shots in one class. Some close-up shots can be misclassified into out-of-field shots in case if only GFPR is analyzed. The proposed method addresses this issue by using edge pixel ratio in combination with grass field pixel ratio. Canny edge detector [32] is employed to compute the edges. EPR is computed by comparing edge pixels to total number of pixels in each frame. Out-of-field/crowd shots contain more edges as compared to close-up shots. For each frame, EPR is compared against a threshold, T_E set to 0.07 in the proposed work after detailed experimentation as described earlier. EPR feature effectively solves the issue of misclassification between the close-up and out-of-field/crowd shots. The current frame is labeled as out-of-field/crowd view if the GFPR meets these criteria ($T_{\mathrm{CS}} < \mathrm{GFPR} < T_{\mathrm{MS}}$) and EPR must either exceed or equivalent to the threshold T_E. Rule-based induction is applied in the same way as earlier to find the most reliable rule that can effectively identify the out-of-field view with maximum accuracy. Two rules are created from the decision tree for out-of-field shot classification. Entropy is computed to find the optimal rule for out-of-field shot classification that is:

$$R_{\mathrm{OS}} = \left\{ \text{Out-of-field Shot}, \quad \mathrm{T}_{\mathrm{CS}} < \mathrm{GFPR} < T_{\mathrm{MS}} \wedge \mathrm{EPR} \geq T_E \right.$$

The input video frame along-with edge pixel distribution and grass field pixel distribution frames for out-of-field shots of various sports genre are shown in Fig. 7. As it can be observed from Fig. 7 that edge pixel distribution is much higher in out-of-field shots, whereas grass field pixel distribution is minimum as compared to other shots.

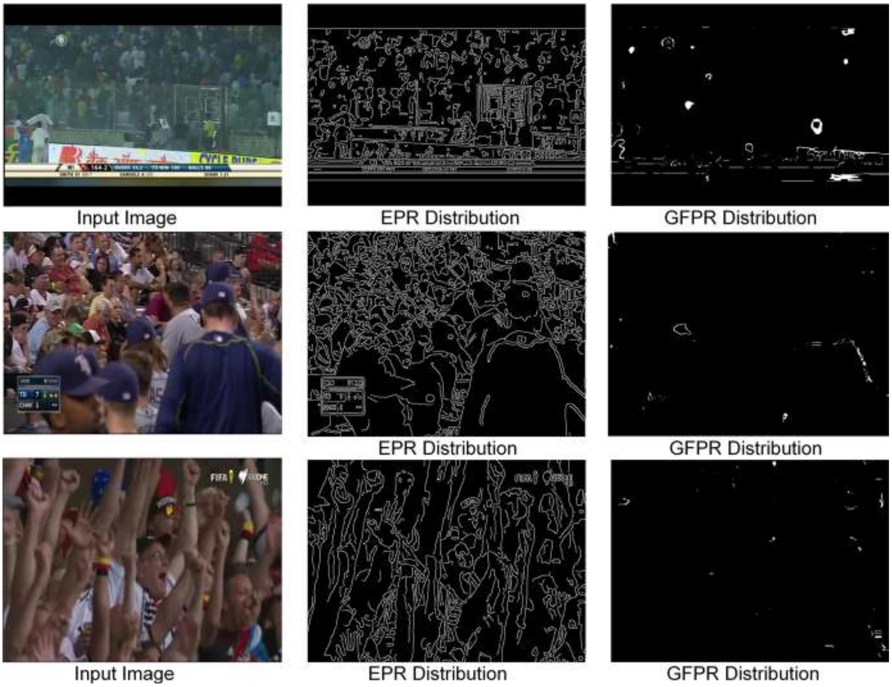


Fig. 7 Row-1: out-of-field shots for cricket, row-2: out-of-field shots for baseball, row-3: out-of-field shots for soccer

3.3.5 Majority voting scheme for shot classification

The proposed shot classification algorithm classifies each frame individually in a shot as either long, medium, close-up, or out-of-field views/shots. A majority voting scheme is employed to assign one shot category to each video shot among the long, medium, close-up, or out-of-field shots based on the following criteria.

$$S_t = \max\{L, M, C, OF\} \tag{5}$$

where L, M, C, OF represents the counters of long, medium, close-up, and out-field frames and S_t is the category of shot.

4 Performance evaluation

Performance of the proposed shot classification method is evaluated on YouTube video dataset of field sports videos. Precision, recall, accuracy, error rate, and F1-score are used for to measure the performance of the proposed framework.

4.1 Dataset

For performance evaluation, we used YouTube video dataset [38] consisting of 50 sports videos of a total duration of 100 h, where video lengths vary from 45 min to 7 h. We selected YouTube dataset [38] to measure the performance of our method as same dataset was employed by existing state-of-the-art [6–8, 17, 21, 22]. Our dataset videos are diverse in terms of sports genre, video length, editing effects, shot types, illumination conditions (i.e., games captured in daylight and artificial-lights), etc.

Each video in the dataset has a frame resolution of 640 × 480 pixels and a frame rate of 25 fps. Videos belong to three sports categories, i.e., Cricket, Soccer, and Baseball. The dataset consists of videos from six major broadcasters, namely ESPN, Star Sports, Ten Sports, Sky Sports, Fox Sports, and Euro Sports. We selected different broadcasters to ensure the diversity in editing effects of the selected videos. The cricket videos contain samples from 2014 One Day International (ODI) series between South Africa and New Zealand, 2006 ODI series between Australia and South Africa, 2014 test series between Australia and Pakistan, 2014 ODI series between South Africa and New Zealand, 2014 (T20) cricket world cup tournament, and 2015 ODI cricket world cup tournament. The soccer videos contain samples from 2014 and 2018 FIFA world cups, and 2016 Euro-cup, and the baseball videos contain samples from 2015 Major League Baseball. Snapshots of the long, medium, close-up, and out-of-field/crowd shots of our dataset for Cricket, Soccer, and Baseball are shown in Fig. 8.

4.2 Experimental results

Objective evaluation criteria are used to measure the effectiveness of the proposed method in terms of shot classification for sports videos. To this end, the proposed

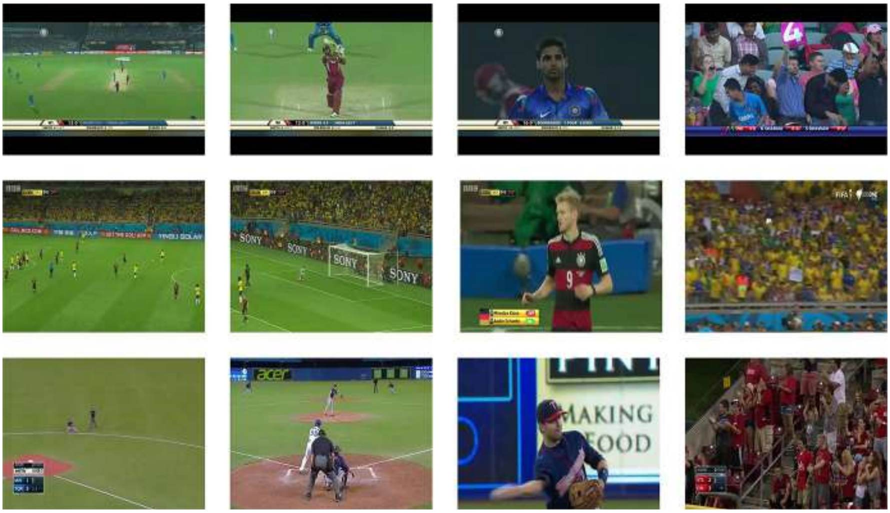


Fig. 8 Snapshots of dataset, row 1: cricket, row 2: soccer, row 3: baseball

method is used to classify each shot of the input sports video into either long, medium, close-up, or out-of-field/crowd shots. We used 60% shots of the dataset for training and remaining 40% for testing purpose. In this section, analysis of the results and various experiments are presented that are designed to evaluate the performance of the proposed method.

4.2.1 Objective evaluation

We designed this experiment to investigate the performance of the proposed decision trees-based framework for accurate classification of long, medium, close-up, and out-of-field shots of the sports videos.

Figure 9 shows the objective evaluation of the proposed shot classification method for Cricket, Soccer, and Baseball videos. The performance of the proposed method in terms of correctly labeling the shot type (i.e., long, medium, close-up, out-of-field/crowd) among the total number of detected shots is remarkable as depicted in Fig. 9. Moreover, the performance of the proposed method in terms of true positive rate of the shot category among the total number of actual shots of that category in the input video is also excellent. The average precision of 91%, 97.65%, and 95.34%, and recall of 92.3%, 98.12%, and 98.43% for Cricket, Soccer, and Baseball indicate the effectiveness of the proposed method for shot classification.

Similarly, it can be observed from Fig. 9 that the ratio of the misdetection for various shot categories is extremely low. Contrary, true detection rate of various shot categories among the total number of shots is outstanding. The proposed method achieves an average error rate of 4.1%, 1.12%, 1.66%, and accuracy rate of 95.9%,

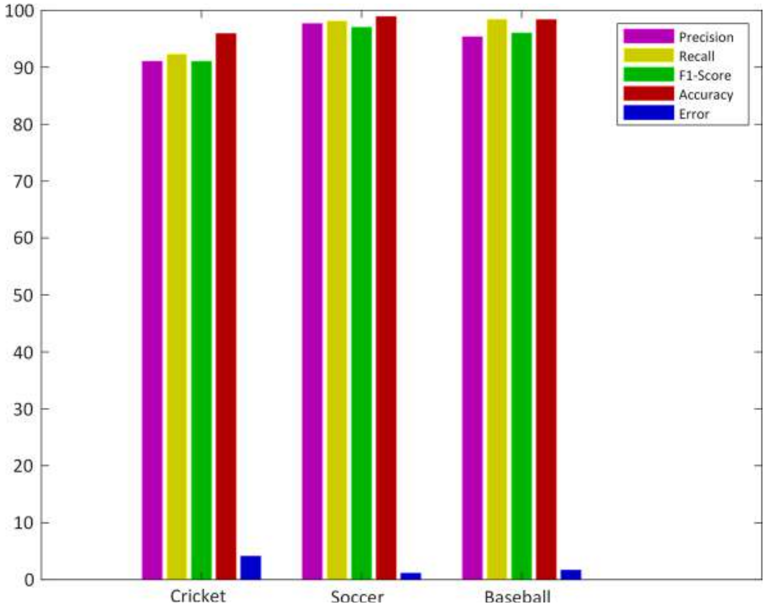


Fig. 9 Detection performance

98.88%, 98.34% for Cricket, Soccer, and Baseball videos. The reported results of error and accuracy rates signify the effectiveness of the proposed method for shot classification.

F-1 score is also computed to measure the test accuracy of the classification performed by the proposed method. F1-score is an effective indicator for performance comparison in case some method has better precision but lower recall over other method. The proposed method achieves the F-1 score of 0.91, 0.97, and 0.96 for Cricket, Soccer, and Baseball, respectively, as shown in Fig. 9. The average F-1 score of 0.94 signifies the effectiveness of the proposed method for better accuracy in terms of shot classification.

From the results presented in Fig. 9, we can observe that the performance of proposed method marginally reduces to some extent for cricket videos as compared to baseball and soccer. This is due to low detection rate of close-up shots in cricket videos.

We also provided the details of the actual and detected shots of each category (i.e., long, medium, close-up, and out-of-field) for sports videos. Table 2 shows the detection rates of the actual and detected shots category for cricket, soccer, and baseball videos. From the results presented in Table 2, we can observe that the proposed method achieves remarkable shot detection performance for field sports videos. It is important to mention that the misdetection rate of close-up shots in cricket videos is high as compared to soccer and baseball videos. This is due to low face detection accuracy in close-up shots of cricket videos where batsman is seen wearing the helmet that blocks the full face exposure of the batsman. In this case, we experience more false negatives in face detection. Upper body detector feature improves the performance of close-up shots detection in this scenario. However, close-up shots where full view of upper body is unavailable and/or contains high density of GFPR, then the proposed method relies more on face detector for close-up shot detection.

4.2.2 ROC curve analysis

In this experiment we evaluated the performance of the proposed method through receiver operating characteristics curve analysis to present the classification performance of the proposed method. The proposed method performs discrete classification as it assigns a class label to each frame as an output. Each discrete classification technique generates an (FPR, TPR) pair which represents a single point in ROC

Table 2 Shots detection performance: actual versus detected shots

Sports genre	LA	LD	MA	MD	CA	CD	OA	OD
Cricket	25,010	28,990	17,000	19,000	36,666	30,334	8000	7000
Soccer	38,007	39,993	35,410	34,590	43,050	41,060	22,500	23,000
Baseball	35,121	37,770	32,100	30,504	29,009	30,022	25,700	24,300

LA, actual long shots; LD, detected long shots; MA, actual medium shots; MD, detected medium shots; CA, actual close-up shots; CD, detected close-up shots; OA, actual out-of-field shots; OD, detected out-of-field shots

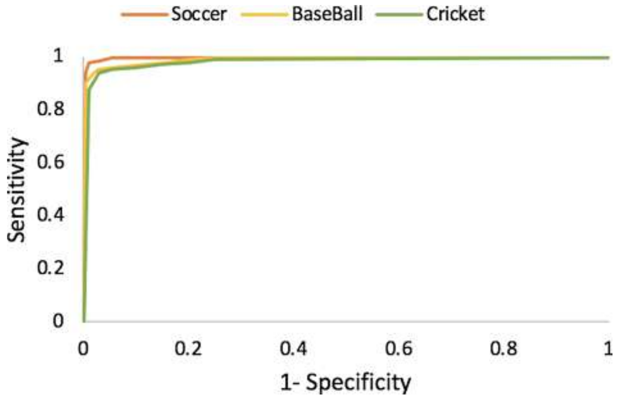


Fig. 10 ROC curve analysis

Table 3 Entropy of selected rules

Selected rules for shots	Entropy
R_{LS}	0.13
R_{MS}	0.17
R_{CS}	0.28
R_{OS}	0.19

space. ROC curves for Soccer, Baseball, and Cricket are shown in Fig. 10. It can be observed from Fig. 10 that the ROC curves of Soccer and Baseball provides excellent classification performance. However, the ROC curve for Cricket shows a slight drop in the performance due to misclassification of close-up shots. This happens in close-up frames of batsman as the helmet blocks the exposure of full face that results to drop the accuracy of face detection method which is also used to classify the close-up shots.

4.2.3 Entropy computation for rules selection

In this experiment, the entropy of various rules in the decision tree is computed to select the most reliable rules for shot classification. Entropy is a measure of impurity of a data sample that tells how well the classes are separated. We need to select those features that significantly reduce the unpredictability in terms of classification for each rule. Entropy is computed using [39] for various rules to filter the least unpredictable rule against each shot (i.e., long, medium, close-up, and out-of-field). We selected one rule against each shot category based on minimum entropy values and results are presented in Table 3.

It can be observed from Table 3 that the computed entropy values of the selected rules for each shot category are quite low. This indicates better predictability of the above chosen rules for long, medium, close-up, and out-of-field shots. Hence, the above selected rules described in Sect. 3.2.3 can reliably be applied to detect the long, medium, close-up, and out-of-field shots for sports videos.

4.2.4 Performance comparison

Performance of the proposed method is compared with existing state-of-the-art to elaborate the effectiveness of the proposed method for shot classification. Detection performance of the proposed and existing shot classification techniques [6–8, 17, 21, 22, 40] in terms of precision and recall is presented in Table 4.

In [7], histogram difference was used in combination of Markov model for shot boundary detection. Dominant color information of the field was used to classify the shot as long, medium, and close-up views. In [22], color and texture features were used to train the SVM to classify the shots into long, medium, and close-up. In [21], radial basis decomposition function (RBF) and Gabor wavelets were used to train the SVM for scene classification of sports videos. In [6], local and full-scene features were used to design a deep learning method for shot classification. In [8], Bayesian network was applied for shot classification of soccer videos. In [17], audio-visual features were used to develop a summarization framework for soccer videos. Grass field color feature based on the specified threshold was used to detect the long shots in soccer videos. In [40], a deep neural network consisting of CNN and RNN was employed for shot classification and key-events detection of soccer videos.

The proposed decision tree framework performs significantly better than [6–8, 17, 21, 22, 40] with two potential benefits. First, our framework is computationally efficient as compared to other state-of-the-art methods. The proposed framework has linear complexity $O(f)$, where f represents the number of features. The depth of our decision tree is six; this operation can be executed in microseconds without any refined implementation. Our method is suitable to analyze the sports videos at real-time. Second, the proposed decision tree architecture provides more interpretability as compared to black-box oriented deep learning models like [6, 40].

The comparative analysis of the proposed and existing methods in terms of precision, recall, and computational complexity of prediction models is provided in Table 4. From the statistics presented in Table 4, we can observe that the proposed method achieves best detection performance and minimum computational cost over comparative methods. Therefore, we conclude from this experiment that the proposed shot classification method can reliably be used to classify the shots of field sports videos.

5 Conclusion and future work

The proposed decision tree framework for shot classification uses low-level features in combination of mid- and high-level features to effectively classify the long, medium, close-up and crowd shots for field sports. Rule-based induction is applied to generate various rules for each shot category. We computed the entropy to select the rules that are least unpredictable in terms of shot classification. Since the baseball and cricket fields consist of both grass field and pitch field, therefore, pitch field

Table 4 Detection performance of shot classification

Shot classification methods	Classifier	Sports genre	Precision	Recall	Computational cost
Fani et al. [6]	CNN	Soccer	90.6	91.3	$O(fn_{l1} + n_{l1}n_{l2} + \cdots)$
Mostafa et al. [7]	SVM KNN	Soccer	90.3	89.9	$O(n_{sv}f)$
Khalig et al. [22]	SVM	Soccer	91	93.8	$O(n_{sv}f)$
Kapela et al. [21]	SVM	Field sports (i.e. cricket, soccer, baseball, etc.)	82.5	84.2	$O(n_{sv}f)$
Raventos et al. [17]	Rule-based thresholding	Soccer	86	80	$O(f)$
Kolekar et al. [8]	Bayesian-based network	Soccer	86	84	$O(f^2)$
Jiang et al. [40]	DNN	Soccer	92.4	87.8	$O(fn_{11} + n_{11}n_{12} + \cdots)$
Proposed method	Decision tree	Soccer, cricket, baseball	94.7	96.2	$O(f)$

pixel ratio feature is used to generate rules for cricket and baseball only. The proposed framework is computationally efficient and suitable to analyze the sports videos at real-time. Additionally, the proposed framework provides more interpretability as compared to black-box oriented deep learning models. The proposed method is robust to illumination conditions, playing fields, variations in camera and motion, shot speed, game genre and structure, etc. Performance of the proposed method is evaluated on three different sports genre (i.e., baseball, cricket, soccer). The average precision of 94.7%, recall of 96.28%, and accuracy of 97.7% illustrates the effectiveness of the proposed framework in terms of shot classification.

At this point, the face detection rate of the proposed method in cricket videos degrades to some extent in case the close-up shots present the batsman wearing helmet. This prevents the full exposure of face of the player and hence, face detection accuracy reduces marginally that ultimately decreases the accuracy of close-up shot classification in cricket videos. Upper body detector feature compensates for the accuracy drop of those close-up shots where upper body view is visible. However, close-up shots containing only the batsman face and/or abundance of grass field pixel ratio rely to some extent on accurate face detection. We are currently planning to investigate this problem in the future.

Acknowledgements This work was supported and funded by the Directorate ASRTD of University of Engineering and Technology-Taxila (UET/ASRTD/RG-1002-3).

References

1. 2018 FIFA World Cup Russia (2019). https://www.fifa.com/worldcup/news/more-than-half-the-world-watched-record-breaking-2018-world-cup. Accessed 16 Aug 2019
2. ICC Mens Cricket World Cup (2019). https://www.icc-cricket.com/media-releases/1277987. Accessed 16 Aug 2019
3. Merler M, Mac K, Joshi D, Nguyen Q, Hammer S, Kent J, Feris R (2018) Automatic curation of sports highlights using multimodal excitement features. IEEE Trans Multimed 21(5):1147–1160
4. Daudpota S, Muhammad A, Baber J (2019) Video genre identification using clustering-based shot detection algorithm. Signal Image Video Process. https://doi.org/10.1007/s11760-019-01488-3
5. Xiong B, Kalantidis Y, Ghadiyaram D, Grauman K (2019) Less is more: learning highlight detection from video duration. In: 2019 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 1258–1267
6. Fani M, Yazdi M, Clausi DA, Wong A (2017) Soccer video structure analysis by parallel feature fusion network and hidden-to-observable transferring Markov model. IEEE Access 5:27322–27336
7. Tavassolipour M, Karimian M, Kasaei S (2014) Event detection and summarization in soccer videos using Bayesian network and copula. IEEE Trans Circuits Syst Video Technol 24(2):291–304
8. Kolekar MH, Sengupta S (2015) Bayesian network-based customized highlight generation for broadcast soccer videos. IEEE Trans Broadcast 61(2):195–209
9. Javed A, Irtaza A, Khaliq Y, Malik H, Mahmood MT (2019) Replay and key-events detection for sports video summarization using confined elliptical local ternary patterns and extreme learning machine. Appl Intell 49(8):2899–2917
10. Javed A, Bajwa KB, Malik H, Irtaza A (2016) An efficient framework for automatic highlights generation from sports videos. IEEE Signal Process Lett 23(7):954–958
11. Javed A, Irtaza A, Malik H, Mahmood MT, Adnan S (2019) Multimodal framework based on audio-visual features for summarisation of cricket videos. IET Image Process 13(4):615–622
12. Dong N, Xing E (2018). Few-shot semantic segmentation with prototype learning. In: 2018 Proceedings of the British Machine Vision Conference (BMVC), p 6

13. Fan J, Zhou S, Siddique MA (2017) Fuzzy color distribution chart-based shot boundary detection. Multimed Tools Appl 76(7):10169–10190
14. Zabih R, Miller J, Mai K (1995) Feature-based algorithms for detecting and classifying scene breaks. Cornell University, New York
15. Ekin A, Tekalp AM, Mehrotra R (2003) Automatic soccer video analysis and summarization. IEEE Trans Image Process 12(7):796–807
16. Tien MC, Chen HT, Chen YW, Hsiao MH, Lee SY (2007) Shot classification of basketball videos and its application in shooting position extraction. In: 2007 Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 1085–1088
17. Raventos A, Quijada R, Torres L, Tarrs F (2015) Automatic summarization of soccer highlights using audio-visual descriptors. SpringerPlus 4(1):301
18. Choros K (2017) Automatic playing field detection and dominant color extraction in sports video shots of different view types. Multimed Netw Inf Syst 506:39–48
19. Wang DH, Tian Q, Gao S, Sung WK (2004) News sports video shot classification with sports play field and motion features. In: 2004 Proceedings of the IEEE Conference on Image Processing (ICIP). IEEE, pp 2247–2250
20. Jiang H, Zhang M (2011) Tennis video shot classification based on support vector machine. In: 2011 Proceedings of IEEE International Conference on Computer Science and Automation Engineering (CSAE). IEEE, pp 757–761
21. Kapela R, McGuinness K, OConnor NE (2017) Real-time field sports scene classification using colour and frequency space decompositions. J Real Time Image Process 13(4):725–737
22. Bagheri-Khaligh A, Raziperchikolaei R, Moghaddam ME (2012) A new method for shot classification in soccer sports video based on SVM classifier. In: 2012 Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). IEEE, pp 109–112
23. Pei SC, Chen F (2003) Semantic scenes detection and classification in sports videos. In: 2003 Proceedings of IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP), pp 210–217
24. Chen SC, Shyu ML, Zhang C, Luo L, Chen M (2003) Detection of soccer goal shots using joint multimedia features and classification rules. In: 2003 Proceedings of ACM International Workshop on Multimedia Data Mining and Knowledge Discovery (MDM/KDD). ACM, p 27
25. Javed A, Bajwa KB, Malik H, Irtaza A, Mahmood MT (2016) A hybrid approach for summarization of cricket videos. In: 2016 Proceedings of IEEE International Conference on Consumer Electronics-Asia (ICCE- Asia). IEEE, pp 1–4
26. Manickam A, Devarasan E, Manogaran G, Priyan MK, Varatharajan R, Hsu CH, Krishnamoorthi R (2018) Score level based latent fingerprint enhancement and matching using SIFT feature. Multimed Tools Appl 78(3):3065–3085
27. Liu G, Wen X, Zheng W, He P (2009) Shot boundary detection and keyframe extraction based on scale invariant feature transform. In: 2009 Proceedings of Eighth IEEE/ACIS International Conference on Computer and Information Science (ICIS). IEEE, pp 1126–1130
28. Stein M, Janetzko H, Lamprecht A, Breitkreutz T, Zimmermann P, Goldlcke B, Schreck T, Andrienko G, Grossniklaus M, Keim DA (2018) Bring it to the pitch: combining video and movement data to enhance team sport analysis. IEEE Trans Vis Comput Graph. 24(1):13–22
29. Jian M, Yin Y, Dong J (2018) Relative flow estimates for shot boundary detection. Pattern Recognit Image Anal 28(1):53–58
30. Deepak CR, Babu RU, Kumar KB, Krishnan CR (2013) Shot boundary detection using color correlogram and Gauge-SURF descriptors. In: 2013 Proceedings of IEEE International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, pp 1–5
31. Coldefy F, Bouthemy P, Betser M, Gravier G (2004) Tennis video abstraction from audio and visual cues. In: 2004 Proceedings of IEEE International Conference on Multimedia Signal Processing (MSP). IEEE, pp 163–166
32. Kim W, Moon SW, Lee J, Nam DW, Jung C (2018) Multiple player tracking in soccer videos: an adaptive multiscale sampling approach. Multimed Syst 24(6):611–623
33. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 886–893
34. Ali G, Iqbal MA, Choi TS (2016) Boosted NNE collections for multicultural facial expression recognition. Pattern Recognit 55:14–27

35. Xiao B, Huigang Z, Zhou J (2014) VHR object detection based on structural feature extraction and query expansion. IEEE Trans Geosci Remote Sens 52(10):6508–6520
36. Wang Z, Wang K, Yang F, Pan S, Han Y (2018) Image segmentation of overlapping leaves based on ChanVese model and Sobel operator. Inf Process Agric 5(1):1–10
37. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: 2001 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 511–518
38. Javed A, Malik H, Bajwa K, Irtaza A (2017) Mahmood MT Data from: replay detection framework for automatic highlights generation from sports videos. Dryad Digital Repository. https://doi.org/10.5061/dryad.5b880
39. Chandrasekar P, Qian K, Shahriar H, Bhattacharya P (2017) Improving the prediction accuracy of decision tree mining with data preprocessing. In: 2017 Proceedings of 41st IEEE Annual Computer Software and Applications Conference (COMPSAC). IEEE, pp 481–484
40. Jiang H, Lu Y, Xue J (2016) Automatic soccer video event detection based on a deep neural network combined CNN and RNN. In: 2016 Proceedings of 28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, pp 490–494

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Ali Javed[1] · Khalid Mahmood Malik[1] · Aun Irtaza[2] · Hafiz Malik[2]

Khalid Mahmood Malik
mahmood@oakland.edu

Aun Irtaza
airtaza@umich.edu

Hafiz Malik
hafiz@umich.edu

[1] Computer Science and Engineering Department, Oakland University, Rochester, USA

[2] Electrical and Computer Engineering Department, University of Michigan-Dearborn, Dearborn, USA