Player Detection and Squad Track from Football Transmit Video

V. Ramya[1]
[1] Assistant Professor
[1]Department of Electronics & Communication Engineering
[1] Saranathan college of Engineering, Trichy-6200 12, Tamilnadu, India

Abstract— *To detect and track the players from a football broadcast video which can be used for player performance analysis. In this paper view classification is performed to distinguish game events from break of play. We propose a novel method that identifies the playfield and detects the players within the playfield in the frames containing game events. The players are then classified into 2 teams based on their jersey colours and tracked using colour features alone from the video. The proposed method provides promising results and although it was tested only on a broadcast video of a football game, it can be extended to other field sports like hockey with very little modifications.*

Keywords— Football, Video Analysis, Shot Classification, Players Detection and Tracking, Team classification

I. INTRODUCTION

Sports videos have attracted lots of attention in automatic video analysis because of its wide viewership and tremendous commercial value. Football is one of the most popular sports across the globe and automated video analysis can either be used to analyse players' performance in a game or for highlights extraction by identifying exciting events such as goals from the video that can be used for broadcasting applications.

Automated player performance analysis can be used to achieve a competitive edge by enabling coaches and scouts to infer the strengths and weaknesses of individual players and teams and focus on strategic planning and tactical decision making and getting every minor detail during the whole game.

A. View Classification

Any football broadcast consists of 3 types of views. They are 1) Longshot or Global view, 2) Close-up or zoom in view and 3) Out of field view. The characteristics of the view types are as follows.

- Global view displays a larger area of the field compared to other shots. It consists of more than 95% of game events. Fig.1 (a) and (b).
- Zoom-in view is a close-up shot of a particular player on the field. Only a small area on the field is covered in zoom-in views and often indicates a break in play. Fig.1 (c).
- Out of field view are the other shots that show the audience and not the field. Similar to close-ups, an out of field shot often indicates a break in the game. Fig.1 (d).

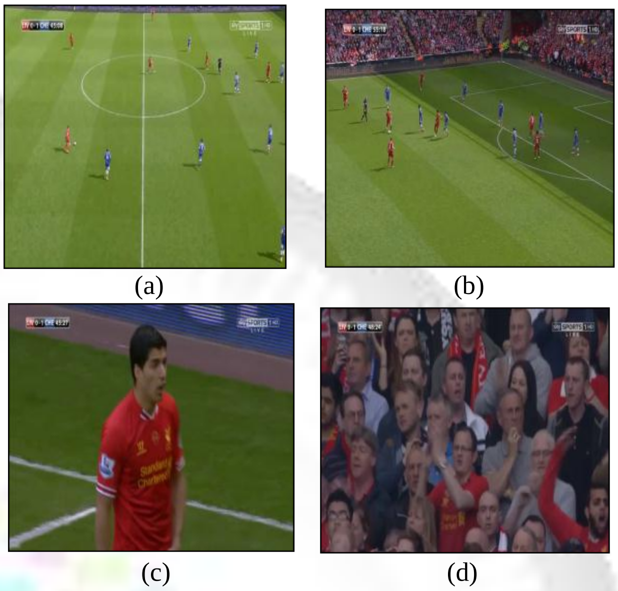


Fig. 1:View Types in Football: (a), (b) Global View, (c) Close-up View, (d) Out of field view.

In any broadcast video, an overwhelming majority of the frames contains those that can be classified as global view. From Fig.1.we infer that colour of grass is the dominant colour and so shot classification can be performed by determining the number of pixels whose colour is the same as the grass colour.

The number of grass pixels is identified for each frame and frames having higher grass pixels are identified as Longshot frames and the rest of the frames are discarded. Zoom-in and out of field views can be neglected when analysing player performance as they mostly contain replays and other break-of play events.

Football games are played at different points in time during the day and to handle different lighting conditions, field conditions (pattern of grass) and shadows, we determine the grass colour of each video by using a set of randomly drawn frames from the initial portion of the video.

50 random frames are chosen and they are converted to HSV (Hue, Saturation and Value) color space. The histogram of the hue component is added up over these 50 frames and the pixel value corresponding to the peak of this cumulative hue histogram is determined as the grass color. The cumulative histogram of 50 random frames is shown in Fig.2. The range of values that is within 10% on either side of the peak is also considered as a grass pixel.

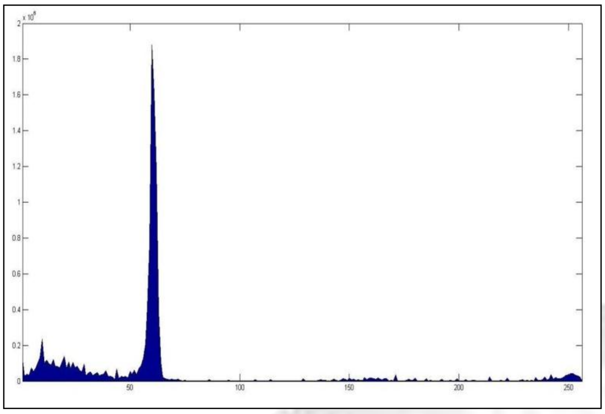


Fig. 2: Cumulative Hue Histogram of 50 random frames

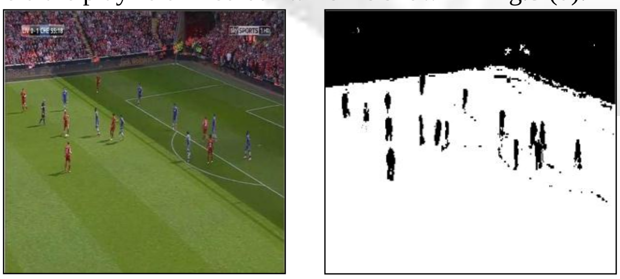
The Grass Ratio (GR) is determined for each frame using the equation (1) and frames having GR greater than a threshold value are classified as longshot frames and used for further processing while the other frames are discarded. A threshold of 0.60 was chosen after trial and error and the results are promising.

$$\text{Grass Ratio} = \frac{\text{Number of Grass pixels}}{\text{Total number of pixels}} \qquad (1)$$

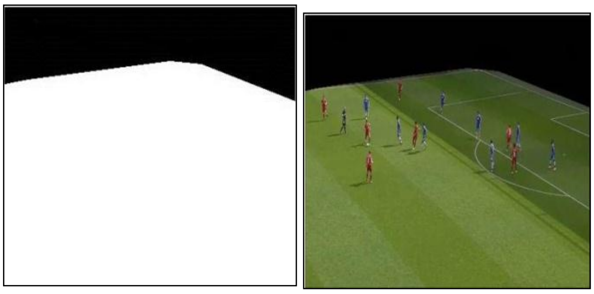
II. PLAYFIELD DETECTION

The main objective behind playfield detection is to avoid misclassification of audience as player regions by limiting the processing area and further increases the robustness of the algorithm. Additionally the detected playfield can be used to estimate the position of the objects of interest in the field. The technique used is again based on the grass colour determined in the view classification module. If the hue value of a pixel in each identified global view frame, Fig.3 (a) falls within the range of values that are already determined as the grass colour, they are set as foreground pixels (white) and the rest of the pixels are set as background pixels (black). The obtained grayscale image is transformed into a binary image by thresholding as shown in Fig.3 (b).

The foreground object having the largest area is determined and computing the Convex Hull of the largest foreground object yields the playfield as shown in Fig.3 (c). This binary image is used as a mask on the original frame to yield the playfield in colour which is shown in Fig.3 (d).transformed into a binary image by thresholding as shown in Fig.3 (b). The foreground object having the largest area is determined and computing the Convex Hull of the largest foreground object yields the playfield as shown in Fig.3 (c). This binary image is used as a mask on the original frame to yield the playfield in colour which is shown in Fig.3 (d).



(a) (b)



(c) (d)

Fig. 3: Playfield Detection (a) Input Frame, (b) Binary Image showing grass as foreground pixels, (c) Convex Hull of the largest foreground object, (d) Detected playfield after applying Binary Mask.

III. PLAYERS DETECTION, TEAM CLASSIFICATION AND TRACKING

The grass colour that has been determined already is subtracted from the detected playfield leaving behind only the players regions. The players are classified into one of the two teams based on jersey colour. In all the previous stages, each frame is processed by representing the pixel values in unsigned 8-bit integer format lying in the range [0, 255] and majority of the processing happened in HSV colour space. Team Classification is done in the RGB colour space and to improve the performance of team classification, colour approximation was performed to reduce the number of colours from 2^{24} to 2^{15}.

Since the two teams are wearing red and blue jerseys, team classification was performed by calculating the ratio of pixel intensity in the red and blue channels respectively using equations (2) and (3).

$$X = \frac{\text{Pixel Intensity in Red Channel}}{\text{Pixel Intensity in Blue Channel}} \qquad (2)$$

$$Y = \frac{\text{Pixel Intensity in Blue Channel}}{\text{Pixel Intensity in Red Channel}} \qquad (3)$$

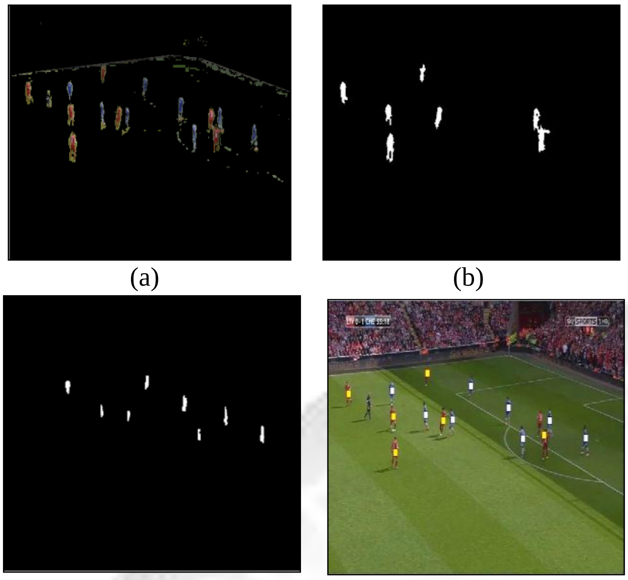
Pixels whose X and Y values were below a threshold value were identified as players belonging to team 1 and 2 respectively and set as foreground pixels in two different binary. The threshold levels were 2 for X and 1.25 for Y respectively.

Basic morphological operations such as closing and opening were performed to take care of the noisy thresholding. To increase the robustness of the method, connected component analysis was performed.

All foreground objects whose area was less than 100 were neglected. The value of 100 was chosen after trial and error.

To avoid false classifications of the scoreboard and broadcaster logo only objects whose orientation was greater than 60^0 were chosen as player regions resulting in Fig. 4(b), (c).

The centroids of all players were determined and the locations of the centroid were superimposed on the original longshot frame with two different colours for the two teams respectively as shown in Fig. 4(d). Extending the above technique to all the frames will result in the players being tracked based on their team.



(a) (b) (c) (d)

Fig. 4: Player Detection, Team Classification and Tracking (a) Player Region after Background Subtraction, (b) Binary Image showing Players from Red Team, (c) Binary Image showing Players from Blue Team, (d) Centroid Superimposed with 2 different Colours

The proposed method to detect footballers, classify them into one of the two teams based on the jersey colour and track them provides promising results. Very small player sizes, camera motion blur, and illumination changes due to strong shadows as well as player occlusions and overlap affect the performance of the tracking algorithm.

IV. Future Work

To completely analyze player performance, ball tracking and estimation of the player position on the football pitch plays an equally important role. Ball tracking poses a major difficulty because of the small size and moreover the ball may be hidden in many frames and fast motion of the ball creates a blur.

Accurately locating the position of the player in the pitch can be done by extracting the lines from the field and finding the corners from the detected playfield. Homography estimation of the players combined with the ball and player tracking provides a complete performance analysis system.

Aggregate trajectory of the ball and players should be determined and local temporal-spatial analysis can be performed to determine key events in the football game such as passing and dribbling.

The proposed method for team classification is based on uniform jersey colour (same colour throughout rather than shirt consisting of stripes of two colors or the shirt and shorts having different colors). A more robust method for team classification has to be implemented to deal with more complex jersey patterns.

V. Results

To compute the accuracy of shot classification, quantifiable metrics to measure accuracy are defined. Precision and Recall are common metrics used for classification.

	Actual Positive	**Actual Negative**
Classified Positive	TP (True Positive)	FP (False Positive)
Classified Negative	FN (False Negative)	TN (True Negative)

Table 1: Precision and Recall metrics

Recall= TP/(TP + FN)

Precision= TP/(TP+FP)

Metrics	Clip	
	Day Game 5 min. segment	Night Game 1 min. segment
True Positive (TP)	6943	1604
False Positive(FP)	1086	302
False Negative(FN)	293	41
Precision	86.47%	84.15%
Recall	95.95%	97.50%

Table 2: Results of View Classification

Metrics	***Team 1 (Red)***	***Team 2 (Blue)***
True Positive (TP)	4253	5632
False Positive(FP)	372	547
False Negative(FN)	355	358
Precision	91.95%	91.14%
Recall	92.29%	94.03%

Table 3: Results of Player Detection and Team Classification

References

[1] Halin, A.A.; Rajeswari, M.; Ramachandram, D., "Shot view classification for playfield-based sports video," Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on , vol., no., pp.410,414, 18-19 Nov. 2009

[2] Kevin Cheng, "Soccer Video Analysis," unpublished. https://stacks.stanford.edu/file/druid:yj296hj2790/Cheng_Soccer_Video_Analysis.pdf

[3] Michael Davis, "Investigation into tracking football players from single viewpoint video sequences", M.S. Dissertation, Bath Univ., England, April 2008.

[4] Guangyu Zhu; Changsheng Xu; Qingming Huang; Yong Rui; Shuqiang Jiang; Wen Gao; Hongxun Yao, "Event Tactic Analysis Based on Broadcast Sports Video," Multimedia, IEEE Transactions on , vol.11, no.1, pp.49,67, Jan. 2009

[5] Ning Zhang, Ling-Yu Duan, Lingfang Li, Qingming Huang, Jun Du, Wen Gao, and Ling Guan. 2012. "A Generic Approach for Systematic Analysis of Sports Videos", ACM Trans. Intell. Syst. Technol. 3, 3, Article 46, May 2012.