Real-Time Identification of Soccer Players Using Sequential Models for Mobile Edge Devices

Aniruddh Shetty
anirudds@andrew.cmu.edu

Kyungzun Rim
krim@andrew.cmu.edu

Nishant Gurunath
ngurunat@andrew.cmu.edu

Abhishek Sawarker
asawarka@andrew.cmu.edu

1 Introduction

Soccer is the most watched and celebrated sport in the world. As the sport grows, more games are being played and it's becoming tactically and physically more demanding for the players. Therefore, there's need to keep track of the movements of the players both to improve their performance and monitor their physical health. Monitoring player movements can potentially help determine training regime, injury prevention, game-readiness and peak performance of the players. In this work, we propose a workflow for efficient tracking of all the soccer players in a match. In particular, we divide the workflow into player detection, player tracking and player identification and, use computer vision algorithms to solve them. In addition to this, soccer fans love to talk about soccer game stats and game plan. In a live game, sometimes it is difficult for the fans to follow the game and especially the players given their distance from the soccer field. The proposed work can also be used to obtain information about each player on fan's smartphone screen. Our goal is to provide succinct information about which player currently has the ball, which team the player belongs to, and general stats and predictions about the game in real-time. We wish to achieve reasonable accuracy for tracking and predictions without using deep neural networks.

2 Related Work

The journal [1] notes several techniques used for this problem in Basketball. They used regressions models in Kalman filter for tracking, CRF models for identification. Some papers have also used DPM for detection or some sliding window + SVM technique. But this is time consuming. Thus we decided to use CV techniques like background extraction and find contours. The paper [2] proposes spatial constellations and player positions to track and identify the players. They use novel resampling and filtering techniques for tracking. There is a lot of ongoing work on finding and tracking the soccer players as mentioned in [3]. They evaluate the computer vision techniques to detect and track players as well as soccer players in the video clips. There are few techniques mentioned in the thesis [2] that contribute to efficient player tracking and detection. [4] also proposes spatial constellations based approach. Here, the style and position of the soccer player is used to identify the player. Also, there are certain commercial applications for solving this problem. We took out initial ideas like sequential models for identification, video processing from [1]. There are a lot of issues that need to be handled to track a player efficiently like occlusion detection, unconstrained environment, changing background, varying players in the view etc. These challenges are listed in [5].

3 Methods

3.1 Detection

This stage takes in the global view image from the previous stage and outputs the detection boxes to the tracking stage. First, we obtain a binary image from input using background extraction (we have calculated the dominant color in the previous stage). Then we apply a morphological close operation on these frames to remove noise present in the crowd. Now we can find contours in this binary cleaned image and apply some area and height thresholds to get the final player detection boxes. We also classify the players into two teams in this stage based on the hsv color histograms of the players. We would know the jersey color of both the teams before the match and we can input the lower and upper hsv bounds of the jersey color.

The problem we are facing currently is that if the players are too close to each other, then they are being clubbed into a single box. We are thinking of applying kmeans on the individual detection box to divide it into multiple boxes depending on the kmeans error.

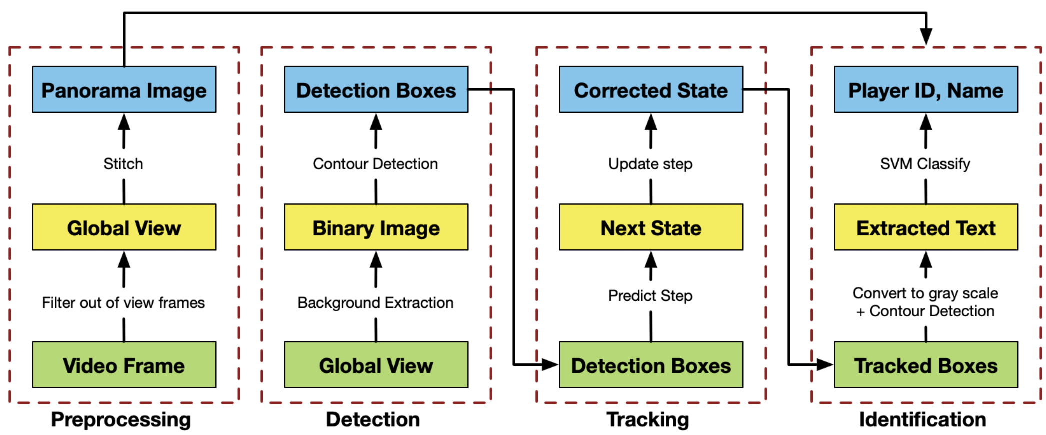


Figure 1: Overall Pipeline

3.2 Tracking

Tracking provides a sequential nature to the model. It predicts the future position of the object based on the current (and past) positions. Tracking can be performed in two ways in terms of the supervision requirements. First, when you only know the initial detection box and predict the futures positions based on this initial detection box as in Lucas-Kanade tracking. Although it performs well in most cases, it has some drawbacks such as when a new object enters the frame or when initial detection is poor. Second, when we use a separate detection algorithm every time-step for supervision as in Kalman Filter tracking. In this scenario, tracking can be assisted using detection and detection too can be corrected using tracking. This enables tracking of new objects and provides robustness to outliers. The figure 3 shows tracking results for a soccer video clip.

We eventually adopted a combined approach of Lucas-Kanade and Kalman Filter tracking, the details of which can be found in the Discussions section. Now, there is another major hurdle to tracking players on a soccer field. The game broadcast happens to keep changing views and camera angles during the match. Then it's very likely to lose track of the players as a result of the discontinuous view. To tackle this problem we define a global view of the field which is essentially a panoramic view of the whole field. The idea is to stitch a panoramic view in the initial segments of the game and then project every frame on this global view so that we can keep track of the absolute position of the players on the field. To perform the stitching we compute a homography between the consecutive frames and warp the images together. We used field line points as our keypoints and used SIFT descriptors to compute homograhy.

3.3 Identification

Once the player is detected and a bounding box is found out over the player, we try to find the jersey number and map it to player name. Identifying player jersey number is a difficult task using conventional methods as the identification highly depends on whether the jersey number is occluded or not, the resolution of view and faithful detection of the player. We tried different techniques to find the jersey number like using HOG features, histogram matching and finally choosing kernel SVM for direct classification. We also tried using OCR on the detected jersey which gave good results.

For our problem as the Jersey number dataset is not readily available, we trained kernel SVM on MNIST dataset first. The test accuracy for 10 class classification on MNIST was about 94%. After pre-processing the detected jersey numbers, namely, converting to grayscale, detecting contours that might have the number, the region of interest was passed through the trained SVM model. The results are show in the figure1. As the SVM was trained on MNIST dataset which has one digit per image, the model is able to classify only one digit from the jersey image. We plan to train the model on multiple images, eg. on SVHN dataset for it to classify multiple digits.

We decided to reinforce our analysis using Jersey number identification

4 Datasets

We are using [6] which contains data of soccer player movements and corresponding videos. The player tracking system provides the player coordinates on the field, their speed, acceleration and force together with an ID and timestamp. The video is captured from the middle of the field using two camera arrays. Kernel SVM for identification was trained on MNIST dataset for jersey number detection and tested on custom test images cropped from video streams of soccer matches. For testing our algorithm, we ran the code on randomly selected soccer match video from youtube. We present few preliminary results on these test videos.

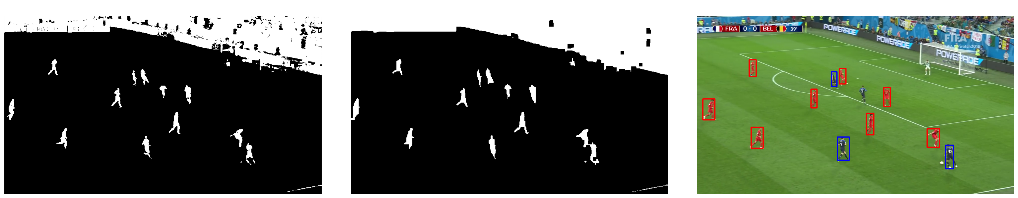
Method	Data	Attributes
Detection & Tracking	Youtube videos	Camera Panning, View Changes
Global View	FIFA game videos	More consistent views
Identification	MNIST	

Table 1: Found Data

5 Results

5.1 Player Detection

In this section, we present the results of our methods. Results for our detection algorithm are shown in figure 2.



(a) Binary image (b) Morphology image (c) The result image

Figure 2: Detected players are marked with bounding boxes in the result image

5.2 Player tracking

The results for tracking algorithm are shown in figure 3

(a) t=0 (b) t=100 (c) t=150

Figure 3: Tracked players are marked with bounding boxes in the result image

5.3 Panorama stitching

The following figure 4 shows our panorama stitching algorithm that merges multiple global view video feeds into a single panoramic image

Figure 4: Panoramic View

In future, we plan to project each video frame onto the panoramic view and then determine a global bird's eye/top view of the field. This is how it's going to look like.

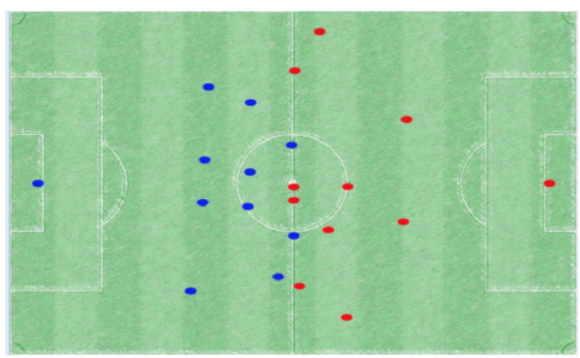


Figure 5: Bird's Eye View

5.4 Evaluation

Since this is not a standard problem, there is no standard dataset on which we can evaluate and compare with baselines. Individual parts of it are pretty standard. But combining all of them in a single pipeline didn't have many research papers. So we took every 10th frame and evaluated using manual annotations and got the below precision and recall.

Method	Precision	Recall	F1
Detection	88.5	71.33	78.9
Tracking	79.04	82.42	80.69

Table 2: Evaluation Metrics

6 Discussion and Analysis

6.1 Detection

There are limitations when we use K-means clustering for the detection part. By using the K-means clustering, we figure out the major color component of the bounding box. As a bounding box includes background color, skin color of the player, and the color of the jersey, we set K as 3. The first limitation is that although the size of bounding box is small, using K-means clustering to every bounding box causes time delay in the detection part. The second limitation is that as we initialize clustering points randomly, the results are not consistent.

6.2 Tracking

As metioned earlier, we decided to move forward with Kalman Filter tracking as it utilizes information from both detection and prediction. We saw pretty decent preliminary result with this model however, it was really hard to come up with a common motion model for all the players who are moving in a very complex fashion. As a result, we decided to adopt a regression based approach to learning motion model for Kalman Filter using information from last 10 time-steps. However, this did not work as expected because whenever detection and existing prediction are far off, resulted in a very random behavior to the motion model and it failed to predict the right states.

Since, we could not determine the right motion model for Kalman Filter, we decided to use Lucas-Kanade tracking to predict the next states for the players and just as in regular Kalman Filter tracking, assisted these prediction using detection. We keep track of the every player on the field and whenever there is detection we try to map detection to the existing tracks. We then make a new prediction for the next time-step based on existing prediction and the new detection. If the detection does not map to any of the tracks we create a new track as this is probably a new player entering the frame.

While stitching the panoramic view, we observed that calculating SIFT features directly on the image is not enough for homography as it majorly detects players as the key points. We realized that the most important feature to stitch a panorama of the field are the field lines. While stitching the images together, we also observed that it is difficult to match key points between a partially stitched image and a new image from the next video frame as the partially stitched image is warped and it becomes difficult to identify the keypoints. So, we resolved to using the original images for homography and kept updating the homography wrt the base image.

6.3 Identification

Identification is a difficult problem as it's difficult to identify players from a distance. We thought of one way to identify the player based on the jersey number. After training linear SVM and Kernel SVM, on MNIST dataset for number classification, we tested our model on randomly picked images of players with their jersey numbers. Kernel SVM gave a good accuracy of around 78% on these images. However, we could not train it for multi-digit recognition due to time and data constraints.

Another approach we tried was using the prior information on player position. This method utilizes the information on player position prior to the start of the match and keeps track of the player during the match. The hypothesis being, once a player is known to occupy a certain position on the field, he's more likely to retain that position for the entire time duration. This also helps in generating a heat map of the player for the entire duration of the game.

6.4 Limitations

We are not able to remove the referee from the detection. He is also being detected along with the players. Also since we are using background subtraction and contour finding, if the color of the player matches that with the playing field, it will pose a serious problem. We are then using color histogram for classifying teams. If both the teams fall in similar hue bin in the histogram, then we can't differentiate them. But this would be rare since soccer teams generally wear contrasting jersey colors. Tracking is limited by the quality of detection, even when there is bad detection it will be tracked. There needs to be a way to distinguish good detections from the bad ones and delete the tracks for the bad detections.

References

[1] W. Lu, J. Ting, J. J. Little, and K. P. Murphy. Learning to track and identify players from broadcast sports videos. volume 35, pages 1704–1716, July 2013.

[2] Murat Durus. Ball tracking and action recognition of soccer players in tv broadcast videos. volume Institut für Informatik Technische Universität München, June 2013.

[3] Huang-Chia Shih. A survey of content-aware video analysis for sports. volume 28, pages 1212 – 1231, January 2017.

[4] Sebastian Gerke. Identifying soccer players using spatial constellation features. 2015.

[5] M.ManafifardH, EbadiH, AbrishamiMoghaddam. A survey on player tracking in soccer videos. volume 159, pages 19 – 46, June 2017.

[6] S. A. Pettersen, D. Johansen, H. Johansen, V. Berg-Johansen, V. R. Gaddam, A. Mortensen, R. Langseth. Soccer video and player position dataset.