Computer Vision impact in Basketball

Computer Vision Course - First Assignment

Riccardo Scilla - 213723
University of Trento
riccardo.scilla@studenti.unitn.it

ABSTRACT

Computer Vision has a strong impact in gathering information for data analytics in sports. This work shows a review of the most common Computer Vision applications and techniques that have been proposed in the recent years to analyze basketball games.

1 INTRODUCTION

The use of Computer Vision is becoming more and more significant in the context of sport analysis, changing the way in which sports are approached by athletes, coaches and spectators. The huge amount of data that can be obtained through efficient algorithms and ad-hoc devices, has significantly impacted the level of audience engagement, creating a strategy for games and the way sports are played today [1].

One of the sports in which Computer vision has been more impactful is basketball given its popularity, especially in the USA: research brings many changes in the ways the sport is seen and played. In this work we will discuss what are the most important applications and the techniques used together with a description of the devices used to gather information from basketball games.

The applications of Computer Vision in basketball are different, starting from the analytics for broadcast which allow to show the position of the player or the ball during the actions or in more detailed replays that can be used by TV presenters or by referees during the game. Research is currently active in the development of more and more advanced devices that allow the 3D reproduction of players during certain actions. Computer vision plays an important role also in training and coaching, helping the players to improve the typical bio-mechanics movements such as shot or pass, and analyzing the ways in which both individual players move together with the overall formation of the team.

2 PLAYER DETECTION AND TRACKING

Automatic player tracking and identification is one of the main aspects in basketball video analysis and also challenging due to several reasons: 1. The appearance of the players is ambiguous because the same team wear the same uniform; 2. Occlusion is frequent also for the limited space where the players typically play; 3. Players have complicated patterns. Due to its importance, a lot of research has been done and few different approaches are described in this section.

A typical approach in this context is doing tracking-by-detection, i.e., run a player detector to locate players in every frame and then associate detections over frames with player tracks.

A work [2] proposes a video tracking system for tracking players in indoor sports using two high quality digital cameras, mounted on the sport hall ceiling. The cameras used in this work present a fisheye distortion of the lens, thus a transformation is applied to correct the video for visualization and to compute the players position in the real-world coordinates.
The procedure explained starts from the process of background detection, creating a background model from the empty field that is updated during the tracking to tackle different lighting conditions. The tracking method used is template matching: the template considered is player's head plus a small part of the shoulder. After locating the best matching position computed with a similarity between the portion of the frame analyzed and the reference template, the player template is updated by replacing the old one with the sub-image of the best match. To track multiple players Voronoi Partitioning [3] is used to identify 10 disjoint regions such that each region contains only one player.

Having an ad-hoc setup of the cameras to detect the court (such as set in hall ceiling) is the optimal solution for having good results in tracking. But this is not always possible, so many techniques have been developed starting from the broadcast videos: this introduce more problems such as an higher probability of having occlusions and a more difficult calibration process.

In this context, a relevant work has been proposed in [4] which develop a robust camera calibration and player tracking system for broadcast basketball videos. The system contains three main components including video-preprocessing module, court detection/camera calibration module, and player tracking module. In the preprocessing steps, irrelevant video clips are filtered according to time-length and dominant color ratio (i.e., the pixels containing mostly the playing field color) so that only those clips with court-view are retained. From these frames, three masks are generated to optimize accuracy and computational cost of the following steps: playing field mask, player mask, player-excluded.

Figure 1: Different masks from court-view frames

The camera calibration module provides a geometric transformation which maps a point in the image coordinates to a point in the real-world coordinates according to the pinhole camera model. In this case, to estimate the homography matrix, the four corners of the free-throw lane are used. These are determined first detecting the dominant court lines, whose parameters are derived applying a RANSAC-based line detector [5]. Second, the quadrangle candidates are generated from the intersection of these dominant court lines. A similar procedure it's described also in [6].
As shown in the paper, this process is quite challenging since the court involves many image noises and also the intersection may be occluded by players. For this reason, a court model fitting is applied taking into consideration specific geometric information of the basketball court and is used to select the homography matrix with the highest fitting score.

Last, in the player tracking phase, the player mask is used to track players in the image coordinates with a modified version of the CamShift-based tracking algorithm [7], to be capable of tracking multiple players in broadcast basketball videos. This procedure starts from different ROIs and finds the local maximum location of the probability distribution; for the successive frame the search window is centered at this maximum and its size is adjusted to tackle player movements. The homography matrix obtained from the camera calibration step is then utilized to obtain player positions. A similar procedure has been described

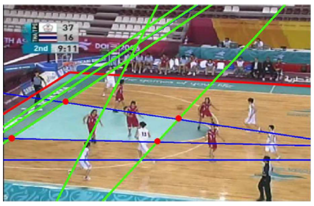

Figure 2: Four calibration points with lines intersection

in [8] which proposed algorithm starts with the court detection using dilation and erosion followed by Canny Edge detector; in this case, Hough transform is used to detect straight lines from the preprocessed frame. To detect players, they used an HOG detector enhanced with a trained SVM, which is typically used for pedestrian and efficiently implemented in the OpenCV library [ref]. This technique is applied also in other works [9]. As shown in the paper, this solution is prone to a high miss rate due to the dynamicity of the scenes. To solve this, a color-based detector and classifier is used: it detects player according to their team and it rules out all the other people that could be wrongly detected by HOG (referees, audience, ...). This classification step starts from the pedestrians detected by the HOG detector, which is more efficient than starting from the entire court as done in [10]. The color detection is made starting applying some thresholding from the HSV color space, which gets rid of the problem of having reflections on the floor. Once detected, players are tracked considering the HOG detector's information obtained in the previous frames: if a player is detected in consecutive frames, its position is simply updated; if instead a player is not detected in the current frame, a color detection is done in the neighborhood of the HOG box of the previous frame to find a correlation. In some situations, such as strong occlusion of players, even the color detection may fail: to handle this, boxes are stored on every single frame and when a player is detected again after some frames, a minimum distance correlator is used to find a correlation between the saved boxes and the new detected box. The player's movements are displayed in a 2D top-view court using affine transformation with an homography map.

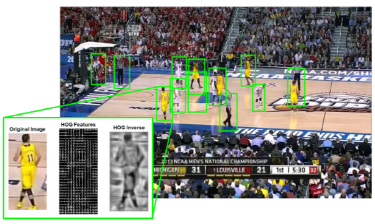

Figure 3: Pedestrian Detection with HOG

More advanced technique used to implement robust player detection involve Deformable Part Model (DPM), as proposed in [11]; the tracking part is done using bi-partite matching where the matching cost is the Euclidean distances between centers of detections and predictive locations of tracks. Once detections are assigned to existing trackers, the state of the players is modeled as linear-Gaussian models and updated using Kalman Filter (or Kalman Prediction if there is no associated detection).

Another feature that has been implemented in this work is the automatic identification of sports players: many possibilities for doing this is using face recognition or retrieving the jersey number and run OCR to recognize them. As we could expect, these solutions are not appropriate for the scope, due to occlusion and the different orientations that players can have when captured from single camera view. For this reason, a more efficient Conditional Random Fields (CRF) is used with several visual features from the entire body of player such as faces, numbers on the jersey, skin or hair colors.

3 BALL TRACKING

Useful information can be obtained not only from tracking players in the court but also from the ball-player interaction. The small size of the ball in relation to the frame size makes ball detection a challenging task in basketball video analysis: it is even more difficult if we consider the collusion and the deformation of the ball-images that are commonly captured with cameras due to the ball's high speed. This task can be handled using computer vision techniques as shown in [12] [13].

The proposed work tries to solve the ball-tracking problem with trajectory interpolation to find out the missing ball locations along the ball flight path. The whole framework provides a near real-time analysis of the ball trajectory without need of multiple camera setup.

The pipeline works as follow:

- Moving Object Segmentation. To segment out the moving objects, it is used a frame differencing method of background subtraction taking three consecutive frames: this method provides robustness in presence of multiple moving objects in the background.
  This is followed by some morphological operations (opening, closing and dilation) to fill gaps and remove noise on the segmented objects.
  A Canny Edge detector is used to discriminate the moving objects from the background.
- Ball Candidate Generation. In this step some filter based on the shape, size and compaction feature are applied to identify ball-like objects.
- Trajectory Processing. From the previous steps we get a number of ball candidates that includes both true ball candidates and non-ball objects. A trajectory-based tracking algorithm is used: first, a candidate distribution plot is generated by plotting the centroid locations of all the ball candidates.
  Next, the trajectories are identified by linking 4 nearest neighbors candidates and predicting the next ball position using a parabolic curve: if the predicted position matches an actual candidate point within a certain *prediction error*, then the trajectory is grown and the parabolic curve is updated using the best-fitting function (obtained with regression).

The best trajectories are then selected considering their total length and a trajectory interpolation method is used to recover the missing ball positions.

A similar version of the same algorithm has been proposed in [14]; additionally they introduced the possibility to calculate the ball throwing angle (considering the coordinates of the ball location in successive frames) and to estimate the ball throwing velocity. This information can be further used to give advices to players and coaches on how to improve the technique in shooting.

Basketball is a sport in which the most important thing is to score more points than the opposite team. Obtaining information on where a team shots the most in the playing field, can be a game-changer factor: the coaches can understand where the opposite players have higher possibility of scoring by shooting and thus enhance the defense strategy, by preventing the opponents from shooting at the locations they stand a good chance of scoring. In this context, [15] proposes a ball tracking algorithm that allow the 3D trajectory reconstruction from broadcast basketball videos, in order to automatically gather the game statistics of shooting locations, i.e., the location where the players shoot the ball and if they score or miss.
The particular challenge faced here is the 2D-to-3D inference which is intrinsically a problem due to the loss of information caused by the 2D projection: to overcome this, the system integrates domain knowledge and physical characteristics of the ball motion.

The process is applied to court-view shots of the broadcast video, so first they are retrieved considering the dominant court color; to improve this step, only the central and central-low sections of the frame are used for the classification.
Camera calibration is performed to provide geometric transformation mapping the positions of the ball and players in the video frames to real-world coordinates or vice versa. However, the points obtained from the court plane are not sufficient to reconstruct the 3D trajectory: in addition to these, some non-coplanar points are taken, specifically the backboard top-border points. As explained in the previous section of player tracking, the points are obtained by intersecting the white lines of the court, in this case extracted applying the Hough transform. A line-structure constraint is applied to reduce false detections, so that only the white pixels with a linear pixel neighbor structure is retained.
Court lines are represented as pair (θ,d), i.e. the angle from the horizontal axis and the distance from the origin: these are used to find the *vanishing point* and the backboard top-border points.
Once all the points are extracted the parameters of camera calibration can be derived to form the matrix which transforms 3D real world coordinate to 2D image coordinate.

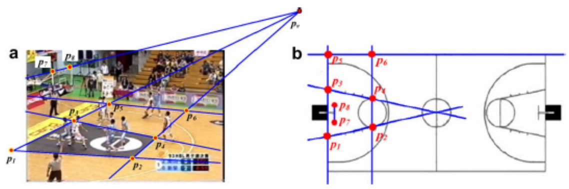

Figure 4: Point-correspondence between 2D frame and court model

The moving objects are extracted using frame differencing method and ball candidates are detected using color feature (HSV compared with histogram of 30 manually segmented images); shape and size properties are used to prune non-ball objects.
Ball trajectories are extracted computing the best-fitting quadratic function, as we have seen in the previous works.
To improve the computational efficiency and accuracy for ball tracking, a in-frame ball velocity constraint is determined from a simplified representation of the dynamic of the long shot: this value tells what is the highest possible speed that the ball can have when shot; this is compared with the velocity of each candidate to decide whether to add it in the trajectory.
With the 2D trajectory extracted and the camera parameters calibrated, the process is able to employ the physical characteristics of ball motion in real world for 3D trajectory reconstruction. The shooting location is estimated by projecting the initial 3D coordinates (x_0, y_0, z_0) onto the court plane.

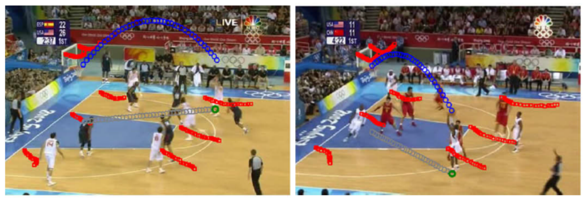

Figure 5: Shooting trajectory (blue), location estimation (green), camera motion (red)

4 SportVU

Patent [16] describes an invention that allow to have a non-intrusive peripheral system and methods to track, identify and capture full motion of athletes, players and other objects (such as balls) on the playing field in real-time. Moreover, it shows how captured body data can be used to generate 3D display of the real sporting event using computer games graphics.

Originally designed for football, this work has been used to develop SportVU: the most popular optical tracking statistic system which is used for analytics by all 30 NBA teams [17]. SportVU uses six cameras positioned every half-court and all synchronized to each other. It is capable of tracking the (x, y) positions of the players along with the (x, y, z) position of the ball 25 times every second. This data is combined with data from the scorer's table, including game events like fouls and turnovers, the game and shot clock, the current score, and other game elements. This raw information is further used to calculate numerous statistics with the aim of improving team performance and providing more information to coaches and referees to improve decision making [18] [19].

5 CONCLUSION

This work has highlighted some of the current applications and techniques for computer vision in basketball, in particular detection and tracking of players and ball. The resulting information that can be obtained are more and more used nowadays to compute detailed analysis that impact the way in which coaches, players and referees take decisions at every category.

REFERENCES

[1] Graham Thomas, Rikke Gade, Thomas B Moeslund, Peter Carr, and Adrian Hilton. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18, 2017.

[2] Emad Monier, Per Wilhelm, and Ulrich Rückert. A computer vision based tracking system for indoor team sports. In *The fourth international conference on intelligent computing and information systems*, 2009.

[3] JP Lewis. Fast normalized cross-correlation, 1995. In *Vision Interface*, volume 2010, pages 120–123, 2010.

[4] Min-Chun Hu, Ming-Hsiu Chang, Ja-Ling Wu, and Lin Chi. Robust camera calibration and player tracking in broadcast basketball video. *IEEE Transactions on Multimedia*, 13(2):266–279, 2010.

[5] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[6] Alexander Embiricos and Gabriel Poon. Single-view 3d reconstruction of basketball scenes. *Retreived on the 16th of September*, 2017.

[7] John G Allen, Richard YD Xu, Jesse S Jin, et al. Object tracking using camshift algorithm and multiple quantized feature spaces. In *ACM International Conference Proceeding Series*, volume 100, pages 3–7. Citeseer, 2004.

[8] Evan Cheshire, Cibele Halasz, and Jose Krause Perin. Player tracking and analysis of basketball plays. In *European Conference of Computer Vision*, 2013.

[9] Kevan Patel Simon Xie, Clive Unger. Basketball player tracking.

[10] Scott Parsons and Jason Rogers. Basketball player tracking and automated analysis. *EE368 final project, Spring*, 2014, 2013.

[11] Wei-Lwun Lu, Jo-Anne Ting, James J Little, and Kevin P Murphy. Learning to track and identify players from broadcast sports videos. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1704–1716, 2013.

[12] Bodhisattwa Chakraborty and Sukadev Meher. A real-time trajectory-based ball detection-and-tracking framework for basketball video. *Journal of Optics*, 42(2):156–170, 2013.

[13] Bodhisattwa Chakraborty and Sukadev Meher. Real-time position estimation and tracking of a basketball. In *2012 IEEE International Conference on Signal Processing, Computing and Control*, pages 1–6. IEEE, 2012.

[14] Bodhisattwa Chakraborty and Sukadev Meher. A trajectory-based ball detection and tracking system with applications to shooting angle and velocity estimation in basketball videos. In *2013 Annual IEEE India Conference (INDICON)*, pages 1–6. IEEE, 2013.

[15] Hua-Tsung Chen, Ming-Chun Tien, Yi-Wen Chen, Wen-Jiin Tsai, and Suh-Yin Lee. Physics-based ball tracking and 3d trajectory reconstruction with applications to shooting location estimation in basketball video. *Journal of Visual Communication and Image Representation*, 20(3):204–216, 2009.

[16] Michael Tamir and Gal Oz. Real-time objects tracking and motion capture in sports events, August 14 2008. US Patent App. 11/909,080.

[17] Colby T Jeffries. Sports analytics with computer vision. 2018.

[18] Mitchell Kates. *Player motion analysis: Automatically classifying nba plays.* PhD thesis, Massachusetts Institute of Technology, 2014.

[19] Basketball team performance- optical tracking.