Single-view 3D reconstruction of basketball scenes
CS 231A Project Proposal

Alexander Embiricos (embirico); Gabriel Poon (gpoon47)

January 31, 2014

1 Introduction

In basketball games, understanding the 3D movement of the players and the ball is essential for the analysis of matches and tactics. As there is considerable demand for this information, proprietary solutions such as SportVU by Stats LLC (Stats, 2013) have been created, but these are expensive and unavailable to the general public. Furthermore, they rely on the use of multiple cameras (6 in the case of SportVU).

For our project, we propose to use commonly available game footage from a single camera to create a frame-by-frame model mapping the locations of the ball and players relative to the court. From our literature review, we are cautiously optimistic that we will be able to do this in near-real-time. For example, Farin et al. (2005) perform camera calibration in 0.06 seconds per frame for tennis and soccer.

While this is an ambitious project with many components, we will describe a breakdown of milestones in section 3 which allows us to adjust the number of steps we accomplish to the encountered difficulties. If we finish earlier than expected, we may extend our project to perform basic rule violation detection, such as travelling violations (walking while holding the ball).

2 Proposed solutions

We can break down the entire project into several sub-projects and will attempt them in the order below. We provide a description of the existing literature as well as our approach for each section of the project.

2.1 Camera Calibration

Camera calibration has been extensively covered in class, and essentially involves the estimation of the camera's intrinsic and extrinsic parameters. Since most of a basketball match is covered from a single camera on the side of the court, we will (probably) ignore footage from all other cameras. The intrinsic parameters of this camera need only be estimated a single time. However, as the

Figure 1: Screenshot of NBA TV Broadcast

camera can both rotate and translate, the extrinsic parameters will need to be estimated for each frame.

Main problem statement: Use known parameters of the basketball court (which we consider to be our world coordinates) such as the markings on the court, to determine the camera's position and orientation in the world using images from a TV broadcast of an NBA game.

There is existing literature on methods of camera calibration for sports videos using courts. Farin et al. (2003) implemented an algorithm for single camera calibration for sports videos using court models. Farin et al. (2005) extended their previous work and implemented a similar procedure with a reported calibration time of 6ms per frame; this is faster than the frame rate of HD videos, which suggests this can be performed in real time.

Our approach here would be to emulate one of the more robust algorithms. While the papers provide extensive descriptions of their procedures, none of them have applied it to NBA basketball. As Figure 1 shows, the lines of the court are in more than one color, and are extensively occluded by objects on and off the court. These are two of the more significant challenges to our problem that the other papers do not encounter.

2.2 Ball tracking and Player tracking

This is the main goal of our project: to map the positions of the ball and the players at a given point in time to a point in the 3D world coordinates as determined by our camera calibration. We will allow the position of the ball to vary in all 3 dimensions, and hope to do the same for the players (although we may have to assume that they are grounded since we have only a single camera and do not know their dimensions without performing significant additional recognition work).

Liu et al. (2006) proposed a method to estimate players' and ball's positions from single-view soccer videos. However, their method of player identification requires substantial areas of empty field around each player, which is not possible for video of NBA basketball. Hu et al. (2011) performed a masking procedure and track players using a Cam-Shift algorithm on videos of basketball games. While they were able to show promising results, their test set is rather small, and it their images are 'cleaner' (i.e. less occlusion between objects). Much of the literature apply different Machine Learning techniques to determine player locations

Our approach to this section would be similar to the previous section - i.e. we will implement one of the more promising algorithms (appears to be Hu et al. (2011) at the moment) and adjust it according to the problems we encounter.

2.3 Action Recognition

Most semantic classification of basketball videos that we have seen has been in terms of overall strategies. (E.g. given the positions and trajectories of each player, what play is being executed) Instead, we will attempt to classify pose-dependent actions at a more basic level, for example identifying when a player takes a step. This could possibly be used as a method for automatic rule violation detection, such as detecting travelling violations.

As such, our approach will be to begin by estimating the pose of a given basketball player as identified in the player tracking component of this project. There are several promising techniques for pose estimation which were mentioned by our TAs, including Sun and Savarese (2011)'s work, which could potentially be used to combine the player recognition above with pose estimation (to be decided, as it may not be suitable for the difficult basketball context, and we would prefer to have separate milestones for these parts). Alternatively, Yang and Ramanan (2011) outline a technique which performs well and is cheap, potentially making it suitable to our efforts to work in real-time.

3 Milestones

We have tried to set our milestones such that if any given step proves too difficult, we can do fewer steps and still result in a "succesfull" project. We thus list our best case scenario below.

1. Camera calibration: 2 weeks → Monday 20140217
   (a) Estimate intrinsic properties for a given match's main camera
   (b) Estimate extrensic properties for each frame of a video
   (c) Optional: optimize to work in real-time
2. Ball & player tracking: 3 weeks → Monday 20140310
   (a) Recognize ball
   (b) Recognize players

3. Action recognition: 1 week → Monday 20140317
   This part is optional and will only be done if we progress faster than expected on other sections.

4. Write report: 1/2 to 1 weeks → Wednesday 20140319

References

Farin, Dirk, Jungong Han, et al. (2005), "Fast camera calibration for the analysis of sport sequences." In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, 4–pp, IEEE.

Farin, Dirk, Susanne Krabbe, Wolfgang Effelsberg, et al. (2003), "Robust camera calibration for sport videos using court models." In *Electronic Imaging 2004*, 80–91, International Society for Optics and Photonics.

Hu, Min-Chun, Ming-Hsiu Chang, Ja-Ling Wu, and Lin Chi (2011), "Robust camera calibration and player tracking in broadcast basketball video." *Multimedia, IEEE Transactions on*, 13, 266–279.

Liu, Yang, Dawei Liang, Qingming Huang, and Wen Gao (2006), "Extracting 3d information from broadcast soccer video." *Image and Vision Computing*, 24, 1146–1162.

Stats (2013), "Sportvu." URL `http://www.stats.com/sportvu/sportvu.asp`.

Sun, Min and Silvio Savarese (2011), "Articulated part-based model for joint object detection and pose estimation." In *Computer Vision (ICCV), 2011 IEEE International Conference on*, 723–730, IEEE.

Yang, Yi and Deva Ramanan (2011), "Articulated pose estimation with flexible mixtures-of-parts." In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 1385–1392, IEEE.