Basketball Player Tracking and Automated Analysis

Scott Parsons and Jason Rogers
Department of Electrical Engineering
Stanford University
Stanford, CA

***Abstract*—In the modern game of professional and collegiate basketball, automated stat tracking, referee rule verification, and video annotation are popular topics. The core aspect of these improvements is player detection and tracking. This paper presents techniques for court segmentation, player detection, team correlation, player tracking, and court to top-down view homography. For a 125 frame sample video, we were able to accurately detect and track the individual players until more complex situations arose, such as players overlapping on the court. In ideal situations, these techniques provided reliable detection and tracking.**

Keywords—player detection; player tracking; homography; basketball; automated sport analysis

I. INTRODUCTION

Automated basketball player detection and tracking has many benefits for both professional and collegiate athletics. Automated statistics could provide teams information about their opposition's plays, formations, and strategy. Real-time image analysis could also enhance the current state of rule verification by removing the referees' human errors from the game. In addition, player tracking could improve video broadcasting by automatically switching to the camera with the best viewing angle, or by focusing on superstars. This would reduce the manual workload, which is primarily how sports broadcasting is done today. This paper provides a Matlab program and techniques to detect each individual player, classify which team he is on, maintain a track on each player, and projects their position onto a top-down view of the basketball court. The Matlab program allows the user to view the raw video and the detected positions side by side for comparison.

II. METHODOLOGY

A. Court Segmentation

First, the video frame is binarized so that the court pixels are the foreground and all other pixels are the background. This step eliminates non-interest areas, such as the crowd. In order to perform this segmentation, a MAP detector is trained by using training masks on the first 10 frames of the video. By isolating the known court pixels over multiple frames, the MAP detector is trained to recognize the average RGB color values of the court pixels. After determining the average RGB values, the MAP detector binary thresholds for pixels within 10% of the averages. This results in a noisy, but clear outline of the court. Morphological operators fill in the black holes and smooth the edges to produce a clean, binary mask (Fig. 1 right). A common player position occurs many times at the edge of the court, so that his body is outside of the court boundaries, such as the top-right Ohio St. player (Fig. 1 left). To mitigate removing these player's bodies from the image, the binary court image is dilated to expand the overall region.

Another major source of noise comes from the scoreboard at the bottom of each frame. This causes major issues since the scores are color coded to match the players jersey colors, which could produce false detections. However, since the scoreboard is in a static position in each frame, it can be filtered out with a binary mask.

Fig 1. The original first frame of the sample video (left) and the final binary court mask of the first frame (right).

B. Player Detection

The next step involves correctly detecting each individual player and classifying which team he is on. Similar to the court segmentation, a MAP detector is trained to recognize the average YCbCr values for each team's jerseys by using a set of training masks [3].

After training the MAP detector, image processing is applied to each frame for player detection. First, the image is multiplied by the binary court mask to remove the noise (Fig. 2). Then the image is binary thresholded based on the expected YCbCr values for each team (Fig. 3). Next, a morphological close operation using a 25x20 rectangular structuring elements increases the size of the largest detections (Fig. 4). A 25x20 structuring element is used because of its resemblance to the players, given that the players are taller than they are wide. Finally, the 10 largest, connected white blobs are detected as player positions for each team. The 10 largest instead of the 5 largest are chosen because the largest blobs do not always correspond to the correct players. By detecting more objects, the tracking function can filter based on track correlation (Fig. 5).

Fig 2. A video frame multiplied by the binary court mask.

Fig 3. The binary court masked video frame after being thresholded based on Ohio St.'s trained YCbCr values.

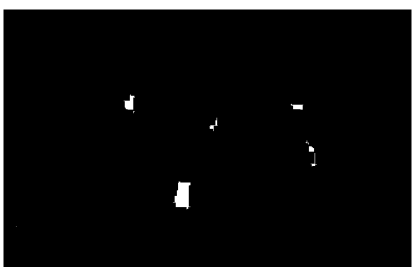

Fig 4. The binary thresholded image after a morphological close operation.

Fig 5. The original first frame of the sample video (left) and the final binary court mask of the first frame (right)

C. *Player Tracking*

After possible players are detected for each team, the system establishes new tracks or correlates the current track to an existing one. For the first frame, the largest 5 connected areas are taken to be the correct positions. The player's (x,y) pixel location is computed as the centroid of the connected area, plus 30 additional pixels in the y direction. This offset pushes the centroid down, which results in a pixel location that is closer to the player's feet, rather than their waist.

After the first 5 player detections in the first frame for each team, the system loops through the subsequent frames and performs track correlation. For each of the 20 detected possible players, it compares the (x,y) location to the 10 established tracks similar to player tracking in [1], [2], and [5]. If the Euclidean distance between the (x,y) locations is within 50 pixels and detections are for the same team, then the tracking function correlates the tracks and updates the current pixel location. If no player detection is found within 50 pixels of an already established track, then it's previous (x,y) location is repeated in it's track structure.

D. *Video Frame Homography*

The final step in the process is to project each player's frame position to their actual position on a top-down view of the court. A key assumption is that the video feed consists of a static camera angle, which does not require the homography matrix to be computed dynamically. By using this assumption, a single 3x3 homography matrix is pre-computed using an affine transform. Each player's (x,y) pixel location is multiplied by the homography matrix, which projects their true position onto the top-down view of the court image (Fig. 6).

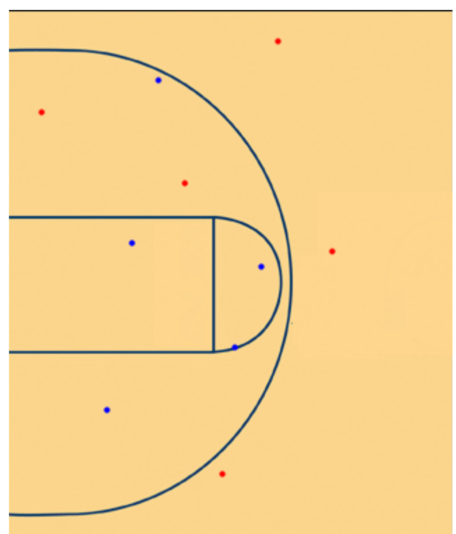

Fig 6. The detected player positions for Ohio St. (red) and Syracuse (blue) after projection using an affine transformation.

III. EXPERIMENTAL RESULTS

As evidenced in Fig. 5, player detection based on centroids gave a rather definitive and reliable detection of the players. Of the 125 frames from the sample video, the player detection algorithm correctly captured 95.6% of players on both teams. As seen in Fig. 8, the yellow track was lost due to players on the same team crossing paths. This decreased the overall accuracy of the system.

Centroid variance was on the order of 30 pixels horizontally and 15 pixels vertically. This translated to approximately 3feet of unfiltered track error after projection to the 2D court (Fig. 7 and Fig. 8).

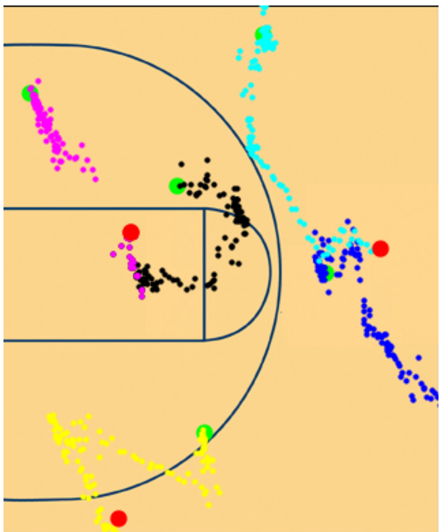

Fig 7. Ohio St. player detections and tracking for all 125 frames of the sample video. Large green and red dots represent the players starting and ending positions, respectively.

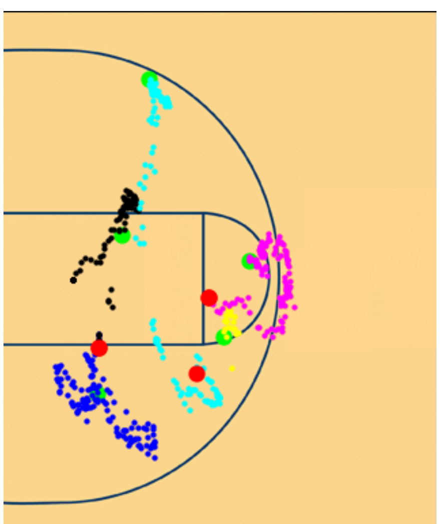

Fig 8. Syracuse player detection and tracking for all 125 frames of the sample video. Large green and red dots represent the players starting and ending positions, respectively.

V. DISCUSSION

For the stationary camera track, player position tracking accuracy was within 3feet of the manually determined track. While this value lacks the fidelity required to achieve the projects stated goals, it serves as a promising proof of concept. With refinement and filtering of the player detection and their tracks, the accuracy can be brought to useful standards.

Because the presented method does not take into count changing camera perspective, it does not hold up against a moving camera. Once the screen pans or zooms on the players, projection rapidly breaks down.

Another area where the method does not yield optimal results is when same team players change direction while one is partially or completely occluding the other. The current algorithm will either merge player tracks or lose one of the players, or it may swap player tracks.

VI. FUTURE WORK

One key feature that is currently missing is automatic court homography determination. There are several approaches which can be implemented including SIFT key point matching of the court and edge and corner detection of the court mask. Both approaches have their own complexity and issues associated with them [1][2].

Player position filtering would greatly improve the fidelity of the system. A Bayesian filter on player positions would remove a lot of the jitter as the centroid detection method [2] [5].

Same team player discrimination using SIFT detection or some other similar recognition technique is necessary to remove track crossing ambiguity and would be critical in determining substitutions and automatic regaining of players when they go of screen or camera angles are switched.

ACKNOWLEDGMENT

The authors would like to thank Professor Bernd Girod, David Chen, Huizhong Chen, and Andrew Araujo for their help and guidance regarding the presented image processing techniques and for their support throughout the course.

REFERENCES

1. Wei-Lwun Lu, Jo-Anne Ting, James J. Little, Kevin P. Murphy, *"Learning to Track and Identify Players from Broadcast Sports Videos,"* IEEE transactions on pattern analysis and machine intelligence, 2011
2. Y. Cai, N. de Freitas, and J. J. Little, "Robust Visual Tracking for Multiple Targets," in ECCV, 2006.
3. Ekin, A.; Tekalp, A.M.; , "Robust dominant color region detection and color-based applications for sports video," Image Processing, International Conference (2003)
4. Min-Chun Hu; Ming-Hsiu Chang, "Robust Camera Calibration and Player Tracking in Broadcast Basketball Video," IEEE Transactions (2011)
5. J. Xing, H. Ai, L. Liu, and S. Lao, "Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling," Image Processing, IEEE Transactions (2011)

Appendix A – Work Breakdown

1. **Code Development**
   A. Court segmentation: Jason
   B. Player detection: Scott
   C. Player tracking: Scott
   D. Homography: Jason

2. **Miscellaneous**
   A.Proposal: Scott/Jason
   B. Presentation PowerPoint: Scott
   C. Presentation Narration: Jason
   D. Final Report: Scott/Jason