Shot Classification of Basketball Videos and its Application in Shooting Position Extraction

Conference Paper *in* Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on · May 2007

DOI: 10.1109/ICASSP.2007.366100 · Source: IEEE Xplore

5 authors, including:

Chen Tsung Huang
South China University of Technology

Yi-Wen Chen
MediaTek

Suh-Yin Lee
National Chiao Tung University

SHOT CLASSIFICATION OF BASKETBALL VIDEOS AND ITS APPLICATION IN SHOOTING POSITION EXTRACTION

Min-Chun Tien, Hua-Tsung Chen, Yi-Wen Chen, Ming-Ho Hsiao, and Suh-Yin Lee

Department of Computer Science and Information Engineering
National Chiao-Tung University, Hsinchu, Taiwan, ROC
{mctien, huatsung, ewchen, mhhsiao, sylee}@csie.nctu.edu.tw

ABSTRACT

In this paper, we propose a system that can automatically segment a basketball video into several clips on the basis of a GOP-based scene change detection method. The length of each clip and the number of dominant color pixels of each frame are used to classify shots into close-up view, medium view, and full court view. Full court view shots are chosen to do advanced analyses such as ball tracking and parameter extracting for the transformation from a 3D real-world court to a 2D image. After that, we map points in the 2D image to the corresponding coordinates in a real-world court by some physical properties of the 3D shooting trajectory, and compute the statistics of all shooting positions. Eventually we can obtain the information about the most possible shooting positions of a professional basketball team, which is useful for opponents to adopt appropriate defense tactics.

Index Terms— scene change detection, dominant color, shot classification, tracking, camera calibration

1. INTRODUCTION

Because of tremendous commercial potentials, research topics on sports videos such as video enhancement, scene classification, indexing summarization, highlight searching, and tactics analysis have been widely studied [1, 2]. Tactics analysis of the basketball video is a tough problem in the domain mentioned above since the background and the camera operations of basketball videos are complicated and fast changing.

Before starting a basketball game, the coach and players have to watch the basketball videos of the opponents and look for their defense rank, offense strategies, offense habitual behavior of the top players and the most possible shooting positions of that team. For human eyes, it is not difficult to grasp the above information. However, it is obviously time-consuming and exhausting to watch a 40 minutes long basketball video if entertainment is not the main concern. Therefore, we propose an approach which can automatically conclude the most possible shooting positions in a game, and provide useful information for the coach. The rest of this paper is organized as follows. Section 2 describes the proposed framework in detail. Section 3 presents the experimental results. Finally, Section 4 concludes our work.

2. THE PROPOSED FRAMEWORK

In this section, we present the framework of our system as depicted in Fig. 1. The system is composed of three main parts: Full Court Shot Retrieval, 2D Ball Trajectory Extraction, and 3D Shooting Location Positioning. Full Court Shot Retrieval utilizes scene change detection to cut a video into clips and classifies each clip as close-up view, medium view, or full court view shot. 2D Ball Trajectory Extraction uses all the full court view shots to locate the ball candidates and to track the 2D ball trajectory. 3D Shooting Location Positioning applies camera calibration to find the relationship between 2D and 3D points. Therefore, we can extract the 3D trajectory of the basketball. Finally the shooting positions are found.

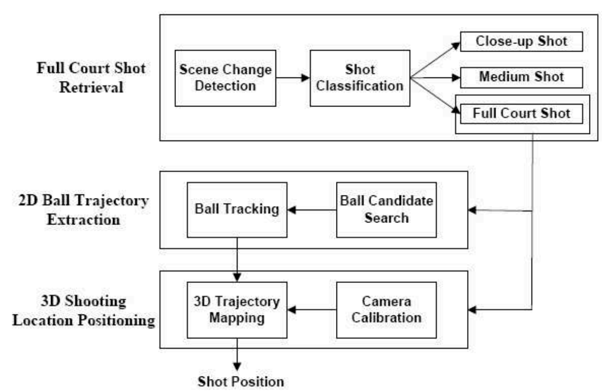


Fig. 1. The framework of the system.

2.1. GOP-based Scene Change Detection

In order to analyze tactics from a basketball video, we have to detect scene changes and divide the video into clips. After that, we classify clips into three kinds of shots and choose the full court view shots to do further processing. Most existed approaches cost a lot of time since they detected scene changes frame by frame. We use a GOP-based method [3] to improve the efficiency of scene change detection. The GOP-based scene change detection approach consists of two steps and the corresponding workflow is shown in Fig. 2. In the first step (inter-GOP scene change detection), possible occurrence of scene change is checked GOP by GOP instead of frame by frame. If a GOP with possible scene change is detected, go to the second step. In the second step (Intra-GOP scene change detection), we check further whether the scene change exists and find the actual frame where the scene change occurs within the GOP.

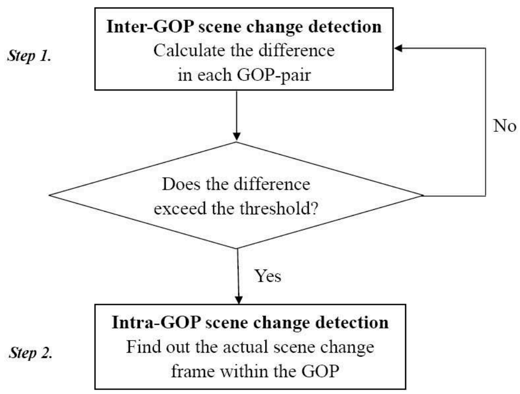


Fig. 2. The workflow of the scene change detection method.

2.2. Shot Classification

Three kinds of basketball video shots, i.e., close-up view, medium view and full court view are defined in advance. Examples of each kind of view shot are shown in Fig. 3. We focus on the full court view shots, which contain most of the information of a game, to do further analyses. The dominant (court) color ratio (DCR)[4] is utilized as the descriptor to extract full court view shots.

(a) Close-up view. (b) Medium View. (c) Full Court View.

Fig. 3. Three kinds of view shot.

A full court view shot usually contains a large number of court pixels, and consequently the distribution of a full court view image is much similar to the color distribution of the court color. With a DCR-based threshold, T_{ratio}, we can filter out close-up view and medium view shots. Since long length clips comprise more information about tactics, we select clips having length larger than another threshold, say L_{min} , to do further analyses.

2.3. Ball Candidate Search

To identify a ball in an image is difficult because the ball is usually small and sometimes moves very fast. The process of ball candidate search is described in Fig. 4.

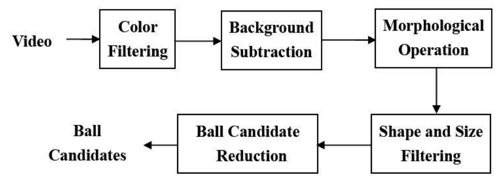


Fig. 4. The process of ball candidate search.

In the color filtering step, color feature is utilized for ball pixel identification. The color of a basketball is not steady owing to the light condition and the viewing angle. In order to study the color distribution of basketball videos, we perform comprehensive experiments. After manually choosing ball blocks from different video sources and calculating their mean values of R, G, B, H, S, and I color components, we observe that most R and G values of a basketball are in the ranges $110 \leq r \leq 175$ and $70 \leq g \leq 135$, respectively. Therefore, we register video blocks having average R and G values in the above basketball color ranges to be possible ball blocks.

Only using the values of R and G is not enough to find out correct ball candidates because of complicated background and noise. Background subtraction is also adopted to select the correct ball candidates. Each possible ball block is compared with the block in the corresponding position of the previous frame. Since basketball usually moves very fast, the ball blocks must have large luminance difference between successive frames. After applying some morphological operations and a region generation algorithm, the region with largest number of connected ball blocks is found.

Many noisy regions rather than the ball region might be detected. Therefore, the area and aspect ratio of the minimum bounding rectangle (MBR) are used as the criteria to locate the possible ball regions. Moreover, we define the ball candidate position as the average coordinate of all ball pixels in the MBR, i.e., $(\frac{1}{n}\sum_{i=1}^{n} Px_i, \frac{1}{n}\sum_{i=1}^{n} Py_i)$, where n is the total number of pixels in the MBR and (Px_i, Py_i) denotes the coordinates of pixel i.

To reduce the number of ball candidates, we perform the Ball-Candidate-Reduction step. Ball-Candidate-Reduction is implemented by examining each ball candidate to see whether there is any other candidate around the search range. The average coordinate of all candidates in the search range is taken as the new candidate position. In this way, many noisy candidates can be deleted.

2.4. Ball Tracking

The foregoing processing helps us to look for possible ball trajectories. We establish a new trajectory list for each ball candidate in the first frame. When a new frame comes into processing, we check if any ball candidate position in the new frame is possible to be the next node in each trajectory list according to the velocity constraint: the velocity of a ball will be in a certain range between two successive frames. Suppose there are two frames, frame i and frame j, the 2D velocity of the ball can be calculated by

$$Velocity_{i\to j} = \frac{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}{t_{i\to j}} \quad (1)$$

where (x_i, y_i) and (x_j, y_j) are the ball candidate positions in frames i and j, and $t_{i\to j}$ denotes the time duration from frame i to frame j.

If there is no candidate added into one list within T_{frame} frames, we check the list to see whether it is a ball trajectory by calculating the variance and distortion of all candidates in the list. Since the ball is usually passed or shot by players, its position will not stay in a small range and its route looks like be a parabola. We estimate the parabola and determine the distortion as the sum of difference between the parabola and each candidate position. For each trajectory list, if the length is long enough, the variance is large, and the distortion is small, it is declared as a possible ball trajectory. The tracking process may result in several 2D trajectories. We will introduce a method to find out the real shooting ball trajectory in section 2.5.

2.5. Camera Calibration

For semantic analysis of sport videos, calculation of camera calibration parameters is a must for converting the positions of a ball and players from 2D image coordinates to 3D real-world coordinates, or vice versa. The geometric transformation which maps points in 3D real-world coordinates to 2D image coordinates can be represented in Eq.(2). We have to find eleven parameters (c_{ij}) to transform an arbitrary 3D world point to a pixel in a 2D image:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \triangleq \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2)$$

where (u, v) denote the 2D image coordinates and (X_c, Y_c, Z_c) stand for the 3D real world coordinates. To calculate the eleven camera parameters, we need at least six non co-plane points whose 2D and 3D coordinates are both known. In court sports like basketball, the marker lines on the court and the backboard boundary can be used to determine the calibration parameters since both the color and length of the marker lines and backboard boundary are determined by official rules. Fig. 5 depicts the correspondences between a 2D court image and a 3D court model and indicates the eight points chosen to calculate camera parameters.

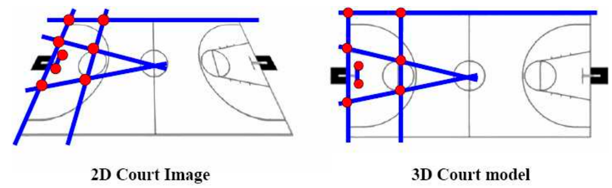


Fig. 5. Line correspondences between 2D court image and 3D court model.

After Applying Hough Transform to extract main white lines of the 2D image [5], the crossings or boundary points of these white lines can be used to calculate the transformation between the 2D image and the 3D real court. Consequently, four points of the backboard are mapped from the 3D real court to the 2D image with the above mentioned transformation. From 2D trajectories obtained in the ball tracking step, we can find a real shooting trajectory by examining whether it passes through the backboard. As illustrated in Fig. 6, the four 2D image points of the backboard are marked as A, B, C and D, which are derived from the 3D real world locations. If the parabola of the 2D trajectory passes through the minimum bounding rectangle of the backboard, it will be a possible shooting trajectory.

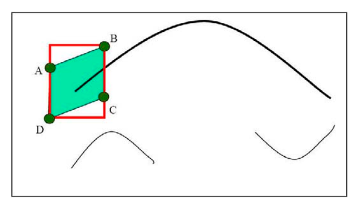


Fig. 6. Extract possible 2D shooting trajectory.

The relationship between each pair of corresponding points in the 2D and the 3D spaces is shown in Eq.(2). Moreover, the 3D ball trajectory should fit the following physical properties:

$$\begin{cases} X_c = x_0 + V_x \cdot t \\ Y_c = y_0 + V_y \cdot t \\ Z_c = z_0 + V_z \cdot t + \frac{1}{2}gt^2 \end{cases} \quad (3)$$

where (x_0, y_0, z_0) is the initial position of the ball in 3D coordinate, (V_x, V_y, V_z) is the velocity of the ball in 3D coordinate, g is acceleration of gravity, and t is the current time. Substituting Eq.(3) into the right hand side of Eq.(2), we obtain

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \cdot \begin{bmatrix} x_0 + V_x \cdot t \\ y_0 + V_y \cdot t \\ z_0 + V_z \cdot t + \frac{1}{2}gt^2 \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (4)$$

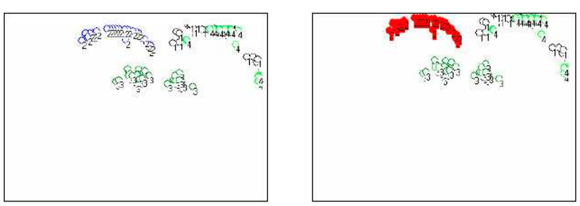
	Close-up		Medium		Full Court	
	Sequence 1	**Sequence 2**	**Sequence 1**	**Sequence 2**	**Sequence 1**	**Sequence 2**
Ground Truth	37	26	27	24	32	21
No.of Miss	1	2	2	2	0	1
No. of False	0	1	1	3	2	1

Table 1. Shot classification result of two testing sequences.

Since the eleven camera calibration parameters and the time for each point on the trajectory are known, we can calculate the six unknowns $(x_0, y_0, z_0, V_x, V_y, V_z)$ of the parabola with three or more arbitrary points on the 2D trajectory. With camera parameters matrix C and six physical parameters of Eq.(3), we can extract a 3D trajectory. As expected, the starting point of the 3D trajectory is treated as a shot position.

3. EXPERIMENTAL RESULTS

We detect scene changes of MPEG encoded testing sequences in the compressed domain. For shot classification and tactic analysis steps, we use AVI sequences and implement the analysis process in the pixel domain. The resolution of all sequences is $360\times$ 240 pixels. Two basketball videos of HBL (High-school Basketball League) in Taiwan are used to test the scene change detection and shot classification algorithm. The first video is of 15 minutes long and contains 96 shots. The second video is of 10 minutes long and contains 71 shots. Table 1 shows the classification results. From Table 1, the accuracy of our shot classification algorithm is about 95.2% (the number of correctly classified shots divided by the number of total shots). The false negative and false positive situations may be caused by the angle of viewing. For instance, if a real full court view shot contains large portion of spectators, the ratio of the court dominant color will be low; consequently, which will result in a wrong classification.

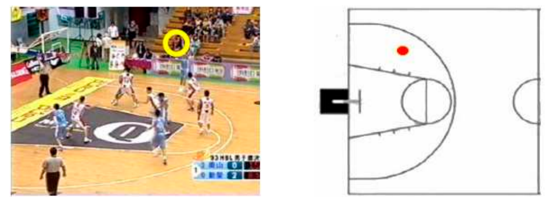


(a) 2D ball positions of all possible trajectories.

(b) Extracted 2D ball positions of the shooting trajectory.

Fig. 7. Result of ball tracking.

Fig. 7 shows an example of ball tracking result: Fig. 7(a) shows the 2D ball positions of all possible trajectories obtained from our ball tracking algorithm, and Fig. 7(b) indicates the extracted 2D ball positions of the shooting trajectory (represented by dark circles). Fig. 8 depicts the result of shooting position extraction on a real-world court: Fig. 8(a) shows the original shooting image, and Fig. 8(b) shows the corresponding extracted shooting position on the court.



(a) Original shooting position in the 2D image (indicated with a hollow circle).

(b) Extracted shooting position on the court model (indicated with a solid circle) .

Fig. 8. Result of shoot position extraction.

4. CONCLUSIONS

We proposed a system that can automatically detect the scene change of basketball videos and classify clips into three kinds of shots. With the full court view shots, we track the ball and define the transformation relationship between 2D image and 3D real-world court model. Finally, the system extracts possible shooting positions. Analyzing tactics in basketball videos is difficult due to the variation of viewing angles, the complexity of backgrounds and the intricacy of court lines. The proposed ball tracking method can be used for any full court view shot no matter whether there is camera motion or not. However, the camera calibration algorithm can only be applied to clips without camera motion. Even though, simulation results demonstrate the applicability of our work.

5. REFERENCES

[1] B. Li, and M. I. Sezan, "Event detection and summarization in sports video," *IEEE Workshop* on 14 Dec. 2001, pp. 132–138.

[2] M. Xu, L.Y. Duan, C. Xu, M. Kankanhalli, and Q. Tian, "Event Detection in Basketball Video Using Multiple Modalities," in *Proceedings IEEE ICICS*, 2003, vol. 3, pp. 1526–1530.

[3] D.Y. Chen, "Towards High-Level Content-Based Video Retrieval and Video Structuring," *Ph.D. Dissertation, National Chiao Tung University, Dep. of CSIE*, 2004.

[4] A. Ekin, and A.M. Tekalp, "Robust Dominant Color Region Detection and Color-based Applications for Sports Video," in *Proceedings IEEE ICIP*, 2003, vol. 1, pp. 21–24.

[5] D. Farin, S. Keabbe, P.H.N.d. With, and W. Effelsberg, "Robust Camera Calibration for Sport Videos Using Court Models," in *SPIE Storage and Retrieval Methods and Application for Multimedia*, 2004, vol. 5307, pp. 80–91.