Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 58 (2015) 225 – 232
Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Object Detection and Tracking based on Trajectory in Broadcast Tennis Video

M. Archana[a], M. Kalaisevi Geetha[b]

[a]*Assistant Professor, Annamalai University, Chidambaram -608 002, Tamilnadu, India*
[b]*Associate Professor, Annamalai University, Chidambaram -608 002, Tamilnadu, India*

Abstract

Ball, player detection and tracking in Broadcast Tennis Video (BTV) is a challenging task in tennis video semantic analysis. Informally, the challenges are due to the camera motion and the other causes such as the small size of the tennis ball and many objects resembles like ball, while the player, the human body along with the tennis racket is not detected completely. In this paper proposed an improved object tracking technique in BTV. In order to track the ball, logical AND operation is applied between the created background and image difference is performed, from that the ball candidates are detected by applying threshold values and dilated. Finally the ball is tracked. Player detection is performed from AND results by finding the biggest blob and filling the whole detected object by removing the small one and the players are tracked based on the contour. The experimental result shows the proposed approach achieved the higher accuracy in object identification, and their tracking. It is achieved a high hit rate and less fail rate for ball tracking while for player tracking is measured by Multiple Object Tracking Precision (MOTP).

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Ball tracking; player tracking; background subtraction; broadcast tennis video; Hit Rate; Fail Rate; Multiple Object Tracking Precision (MOTP).

1. Introduction

An automatic analysis of sports video is an interesting area which attracts many research attentions for several applications, Sports video contains rich audio and video information within a well-organized structure. Owing to increase in the growth of videos on broadcast and internet, there is a need to access semantic events among the full-length videos arises [1]. Instead of accessing the whole lengthy voluminous videos, access of highlights and skipping the less interesting parts of the videos will save not only the viewers time but also the cost. To attract the users the content based views are developed based on their own preferences. Consider the tennis video, the moving object is

* Corresponding author. Tel.: +91 9842897771.
E-mail address: archana.aucse@gmail.com and archana.m.15074@annamalaiuniversity.au

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.060

ball and player around the ground region. Tracking of ball and player in tennis video faces many challenges [2], since the ball focus the attention of viewers and players follows the ball movement in tennis. The main aim of tracking is event detection in Broadcast Tennis Video (BTV) based on the tracking results of ball and player. In tennis the court length is 78 feet, the width is 27 feet for Singles and 36 feet for Doubles.

To track the ball and player some of the challenges are,

1. The target object i.e., the size of the ball is too small in different angles and views. Based on various lighting conditions, the ball may not be visible.
2. Tracking the ball based on trajectory, is little bit complicated because of fast ball and camera movement.
3. The player tracking is considerably easy but dynamic background is a challenge.
4. The size of the players also changes at different angles.

In this context of tracking a tennis ball, noise is a big issue because of the ball size. Due to the quality of the frame, noise appears very frequently among images, which interferes with the process of object detection. Traditional background subtraction approach is not capable of eliminating the majority of noise and they usually require additional operations. A modified background subtraction approach is applied to overcome the limitations due to the bad quality of the captured images [3], [4]. The rest of the paper is organized as follows. Section 2 reviews the related works. The problem analysis is discussed in section 3. The methodology of work and proposed model is discussed in section 4. Experimental results are briefly discussed in section 5 and followed by performance measures in section 6. Finally conclusion and future work are presented in section 7.

2. Related Work

F.Yan, W.Christmas et al. [5] have proposed a tennis ball tracking algorithm for low quality video recorded with a single camera. In this paper a particle filter with improved sampling efficiency is used to track the tennis candidates. Smoothing and observation origin identification are then used to refine the trajectory, to give higher tracking accuracy. Xinguo Yu et al. [8] have proposed a about improving trajectory-based ball detection and tracking algorithm in tennis videos, this algorithm can obtain not only higher accuracy in ball identification, but also ball landing frames and positions with the aid of homography. Yang Wang et al.[11] have proposed about the detection and tracking of player in BTV, this is obtained by support vector classification and court segmentation from that accurate player area is founded. Based on this particle filter tracking of small particles is used to improve the performance of this method.

F.Yan et al. [14] have proposed a ball tracking using automatic annotation of tennis match for low quality video recorded with a single camera. A particle filter is used to track the tennis candidates to achieve better results. It shows the higher accuracy for tracking, while smoothing and observation of origin identification is used to refine the trajectory and it is suitable for tennis annotation. Christmas et al. [9] have proposed about the automatic annotation of tennis video. In this method, proposed an automatic analysis of tennis video which extent the video into individual video shots with accuracy. In some locations, the ball events are wrongly classified and these situations are difficult to realize even by eye. HMM is used for interpreting the play rules in the high level module, which is currently using as input hard decisions made in other modules and for soft decisions, with confidential information.

Min-Yuan Fang et al. [6] have proposed a player tracking of tennis videos using adaptive kalman filter, the parameters of this filter is adjusted based on the detection of players. This improves the tracking accuracy by corrects the detection errors. The success rate of player tracking is well done in singles as well as for double matches. Keni Bernardin and Rainer Stiefelhagen [10] have proposed evaluating metrics i.e., Multiple Object Tracking Performance for any object tracking and also discussed about simultaneous tracking of multiple persons in real-world environments

3. Problem Analysis

Detection and tracking of the ball and player in tennis video face many problems. This section analyses the problems related to tracking. For the purpose of tracking in BTV, the whole tennis court is partitioned into upper and lower half court due to players distinct size. The camera view makes the objects in the upper half much smaller than the lower half and also the background color in the lower half doesn't vary in a random manner [12]. So the tracking of the lower half player isn't a difficult task. On the contrast the detection and tracking of the upper half player are a real challenge. This is because the upper half background varies in a random manner and possesses additional objects

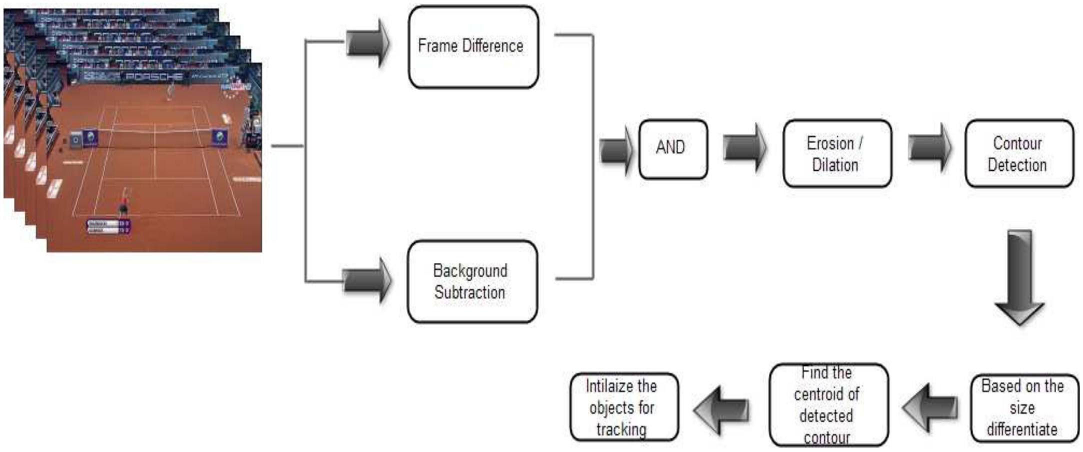

Fig. 1. Illustration of the proposed architecture

such as advertisements, referee and ground staff are often mixed with the background. Hence, it is difficult to generate a stable background and the small size of the upper half player results in poor detection. Though there is a camera behind each baseline, the tracking of upper half player is a hurdle, irrespective of the camera position [13].

4. Methodology

Ball detection is achieved by frame differencing between the current and consecutive images. The results are then verified against the size and shape parameters. The region is close to the expected position, which is chosen in the case of multiple detections. But this technique has the problem of double detections (i.e.) more than one ball candidates appear [7]. To solve this problem first finds the regions in the current image that lies in the expected intensity range for the ball. Then perform a logical AND operation between the result generated from the image difference and the region of background images.

4.1. Proposed Model

In this proposed model, the given images are smoothened, then accumulate the background images and finally create an average background model. After creating this model, images difference is performed between current and next frame. Then logical AND operation is done in the created background image and obtain image difference result. Finally, the ball and the player candidates are detected as illustrated in Fig. 1. Based on the size of the detected contour differentiate the ball and player candidates, after that find the centroid of the detected contour and using the centroid follows the motion of the ball and players separately.

4.2. Smoothing the image

To smoothen the image, median filter is applied, which is a commonly used technique for reducing small noise in an image. Small noises normally appearing very distinct and its gray values are quite different from its neighbors so using this technique eliminated the noise by changing its gray value to the median of neighboring pixel values.

4.3. Background model

The aim of the background model creation is to develop a standard background, in the BTV does not possess a fully static background because of the camera calibration. In this technique, first a background model is created from a

collection of background images. For each queried image, the edges of foreground objects are detected by background subtraction. The tennis ball is identified by shape, aspect ratio and compactness. This technique is analyzed as follows:
1. Given a number of background images, a background model is created.
2. For each queried image:
a. An edge image is created.
b. The edges of the tennis ball are segmented from the result, our approach is based on shape recognition.
c. Area of the tennis ball is then dilated by subjecting to the size and aspect ratio. Finally, the location of the tennis ball is detected.

4.4. Frame difference

Frame Difference (FD) is the technique to find the tennis ball candidates by considering the difference between current and next frames. The reason of choosing the FD is due to the fact the tennis ball is normally fast moving. It will occupy an entirely different set of pixels in consecutive frames, where slow moving objects will have overlaps. In order to estimate the ball intensity levels some of the difficulties are lighting, shadows and distance variations. To get better results, combine the image difference with background subtraction. Motion information is extracted from the video sequence by pixel-wise differencing of consecutive frames. Motion information T_k or difference image is calculated by using Eq. 1.

$$T_k(i, j) = \begin{cases} 1, & \text{if } D_k(i, j) > t; \\ 0, & \text{Otherwise;} \end{cases} \tag{1}$$

Where D_k is the difference image and t is time interval, calculated using Eq. 2.

$$\begin{aligned} &D_k(i, j) = |I_k(i, j) - I_{k+1}(i, j)| \\ &1 \le i \le w, 1 \le j \le h \end{aligned} \tag{2}$$

4.5. Logical AND operation

The logical AND operation is performed between the image obtained from background subtraction and image difference. While applying AND operation the commonly presented objects between these two images are obtained. Finally obtained result is threshold and dilated, based on the compactness, aspect ratio and size, the remaining contour is identified and detected as a ball in the upcoming frames.

4.6. Improved Ball Tracking

In order to detect the ball candidates, given the current image A and the previous image B, ball candidates from A are detected as follows:
1. The image difference between A and B is performed and the result is denoted as C.
2. Perform a background subtraction of image A with the created background model and the result is denoted as D.
3. Perform the logical AND operations between C and D to get the candidates ball.

After detecting the ball candidates, to track the ball centroid is computed for the detected contour then this is plotted in new tracking window, where the window size is as its original frame size. Following the each and every contour is plotted in a window. In Fig. 2. shows the bad visibility of the tennis ball, here the red circle shows the ball which are not visible due to ball‘s high speed movements, where the centroid of the ball candidates are tracked some of the misplaced tracking is also possible due to blurring.

Fig. 2. Bad visibility of the tennis ball

4.7. Improved Player Tracking

The player is detected using the same approach as described in our first technique in section 4.5. The results of AND operation are further pre-processed to remove small blobs. The largest blob is detected which belongs to parts of the human body. The other part of the human body is also detected based on the flood fill technique to reconstruct the player. The following steps are performed,
1. Remove the smaller blobs from the AND results.
2. Detect the biggest blob and it is splitted, which is a part of the human body.
3. To construct the whole human body, find the contour near the big one.
4. To detect a small blob around the biggest one is considered as one by using flood filling techniques.
5. Apply flood fill techniques, it fills the complete region and constructed the region as the human body.
6. Reconstruct the human body, it is also considered the tennis racket in together as player.

To track the player moving along with a camera to follow the action, divide the frame as upper and lower. The upper layer is a challenging task of tracking because of the upper players size is too small and for lower layer is tracked using existing technique such as background subtraction. Fig. 3. shows the player position and the centroid for tracking

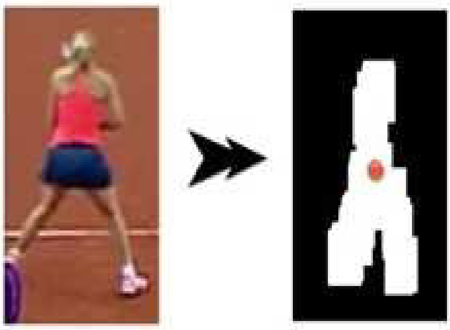

Fig. 3. Detection of player centroid to track

5. Experimental Results

Experiments were carried out on BTV from the different tennis tournaments. A BTV is composed of various shots such as play shots, crowd, advertisement and break.

Fig. 4. Example video frames

To evaluate the performance of the proposed methods, experiments were carried out on video sequence from two tennis tournament such as Australian open and Wimbledon open as seen in Fig 4.The top row shows the video sequences with Australian open and the bottom row with Wimbledon open. The video sequence included a total of 125 shots of (3 - 5 seconds). The experimental data set is in avi format with 1280 x 720 resolution at 25 frames/seconds. In fig 5 shows the position and the trajectory of the ball corresponding to the frame number. Here, the yellow circle denotes the position of the ball and the green line denotes the trajectory of the ball.

Fig. 5. Ball trajectory tracking

Fig.6. shows the position and the trajectory of the player corresponding to the frame number. Here, the blue circle denotes the position of the players and yellow line denotes the trajectory of the lower half player while green line denotes the trajectory of the upper half player.

Fig. 6. Player trajectory tracking

6. Performance Measure

The performance of the proposed model for ball tracking is measured by plotting the ground truth values. The ground truth of the ball was plotted in the original frame. The results were obtained and compared with the detected regions of both the ball with 2 different BTV. Player tracking is measured by means of MOTP (Multiple Object Tracking Precision) because of two players.

Table 1. Rate of success for ball tracking

Tennis Match	Total Frames	Ball Occurred Frames	Hit Rate (%)	Fail Rate (%)
Australian Open	1000	750	90.64	9.36
Wimbledon Open	950	800	90.32	9.36

Table 2. Rate of success for player tracking

Tennis Match	Total Frames	Player Occurred Frames	Player1 Upper Layer (MOTP)	Player2 Lower Layer (MOTP)	Over all (%)
Australian Open	1000	950	85.96%	91.23%	88.59
Wimbledon Open	900	830	84.52%	92.15%	88.33

To find the rate of success and location error with the reference of object center, and the method applied is used to compute the rate of success and score is defined as given below in Eq. 3.

$$score = \frac{area(ROI_D \bigcap ROI_G)}{area(ROI_D \bigcup ROI_G)} \tag{3}$$

Where, ROI_D is the detected bounding box and ROI_G is the ground truth bounding box. The rate of success is computed in all frames, when this score is greater than 0.5 in each frame, then it is considered as a success as shown in Table 1. This work, consider 1000 frames the ball is seen in frames only 60% of frames. The ball detection varies due to some challenges like various lighting conditions, different angles and various shapes due to camera position.

To find the tracking accuracy of the player, here MOTP metrics are used to measure. To track more than one object MOTP is only measuring metrics, here the distance c_t be the number of matches found in time t. For each of these matches, calculate the distance d_t^i between the object o_i and its corresponding hypothesis is defined as given below in Eq.4.

$$MOTP = \frac{\sum_i d^i{}_t}{\sum_t c_t} \tag{4}$$

7. Conclusion

In this proposed model, an improved ball and player tracking for BTV, besides varying camera motion causes as the presence of many balls-like objects and the small size of the tennis ball. It is not only increases the accuracy in identifying the ball, but also improves the accuracy in determining the ball projection position. In addition, it detects the ball landing frames and landing positions based on the accurate ball tracking, whereas the player detection and tracking constructed the human body along with a tennis racket. The result shows, model that this is able to precisely classify the ball tracking in Australian Open about 90.64% and Wimbledon Open of 90.32%. Then for Australian Open Player 1 tracking i.e., Upper Layer of 85.96% and in Player 2 tracking Lower Layer of 91.23% over all achieved player tracking accuracy is 88.59% for and Wimbledon Open also Player 1 tracking i.e., Upper Layer 84.52% in Player 2 tracking Lower Layer of 92.15% over all achieved player tracking accuracy is 88.33% were founded by MOTP metrics. The tracking of the detected objects such as ball and a player is improved by using this proposed approach and it will be investigated in the future.

References

1. Wang, J. R., & Parameswaran N. Survey of sports video analysis research issues and applications. In Proceedings of the Pan-Sydney area workshop on *Visual information processing* Australian Computer Society, Inc 2004; 87-90.
2. Zhu, G., Xu, C., Huang, Q., Gao, W., & Xing, L. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. *In Proceedings of the 14th annual ACM international conference on Multimedia ACM* 2006; 431-440.
3. Jiang, Y. C., Lai, K. T., Hsieh, C. H., & Lai, M. F. Player detection and tracking in broadcast tennis video. Advances in Image and Video Technology *Springer Berlin Heidelberg* 2009; 759-770.
4. Bloom, T., & Bradley, A. P. Player tracking and stroke recognition in tennis video. *APRS Workshop on Digital Image Computing (WDIC03) the University of Queensland* 2003; **1**: 93-97.
5. Yan, F., Christmas, W., & Kittler, J. Ball Tracking for Tennis Video Annotation. *Computer Vision in Sports, Springer International Publishing* 2014; 25-45.
6. Min-Yuan Fang, Chi-Kao Chang, Nai-Chung Yang, Chung-Ming Kuo, Shih-Ku Guang. Robust Player Tracking for Broadcast Tennis Videos with Adaptive Kalman Filtering. *Journal of Information Hiding and Multimedia Signal Processing* 2014;**5(2)**:242-262.
7. Cross R.: *The kick serve in tennis*. Sport Technology 2011; 4: 19-28.
8. Yu, X., Sim, C.H., Wang, J.R., Cheong, L.F. A trajectory-based ball detection and tracking algorithm in broadcast tennis video.*In IEEE International Conference on Image Processing* 2004;**2**:1049-1052.
9. Sun, Y., Lee, K.: *Video compression coding effects on motion detection*. Information technology 2013; 02: 12-129.
10. Keni, B., & Rainer, S.: Evaluating multiple object tracking performance: the CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, Cross R.: Impact of sports balls with striking implements. Sport Eng 2008; **17(1)** :3-22.
11. Goodwill, S.R., Capel-Davies, J., Haake, S.J., Miller, S.: Ball spin generation by elite players during match play. *In: Miller, S., Capel-Davies, J. (eds.) Tennis science and technology* International Tennis Federation, London 2007;**3**; 349-356.
12. Connaghan, D., Moran, K., & OConnor, N. E. An automatic visual analysis system for tennis.*Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology* 2013; **227(4)**: 273-288.
13. Rusdorf, S., & Brunnett, G.: Real time tracking of high speed movements in the context of a table tennis application. *In Proceedings of the ACM symposium on Virtual reality software and technology ACM* 2005; 192-200.
14. Yan, F., Christmas, W., & Kittler, J.: All pairs shortest path formulation for multiple object tracking with application to tennis video analysis. In British Machine.