# Video analysis of a snooker footage based on a kinematic model

Aysylu Gabdulkhakova, Walter G. Kropatsch

Vienna University of Technology
Pattern Recognition and Image Processing Group
{aysylu,krw}@prip.tuwien.ac.at

**Kurzfassung.** Taking an inspiration from psychological studies of visual attention, the contribution of this paper lies in prediction of the critical points of the trajectory using the structure of a scene and physical motion model. On one side, we present our approach for video analysis that differs from traditional tracking techniques by predicting future states of the moving object rather than its next consecutive position using the physically-based motion functionality. On the other side, we propose to use the structure of the scene, which contains the information about the obstacles and space limits, for discovering the critical points of the trajectory. As a proof of concept we developed the use case application for analysing snooker footage.

**Schlüsselwörter:** video analysis, tracking, snooker, visual attention, spatio-temporal prediction

## 1 Introduction

Tracking covers a vast number of applications in computer vision: from media production and augmented reality to robotics and unmanned vehicles [7]. By definition[7] it is a process which aims at defining the position of the target in subsequent video frames. For the purpose of video understanding and summarization storing the entire track of the objects is semantically and computationally redundant. Moreover, high frequency temporal sampling, high resolution data and contamination (i.e occlusions, distractors, illumination variations) makes tracking a challenging task to robustly perform in real time. Thus, we inquired into a question of reducing the amount of processed data by predicting the future locations of the analysed object and select then semantically meaningful moments which require further detailed analysis, such as recognition.

We derived our idea from psychological studies of human vision[12], [14]. Selective visual attention is a mechanism that aims at reducing the amount of visual data. The aim is to filter irrelevant parts of the scene and focus cognitive processes on the most important object at a given time slot. It is accomplished based on either spatial location or object features. It is worth mentioning that we are not aiming at modelling this phenomena. We use its main principles, in order to argue in favour of neglecting redundant data.

As a descriptor motion is assumed to be an inherent property of an object taken apart from environment. Though, motion parameters are dependant not only on the characteristics of the object, but also are restricted by the *structure of the scene*. The correct motion model enables to predict the trajectory, whereas structure of the scene defines the conditions when the trajectory deviates. *Critical point* of the trajectory - is a point of the trajectory where the direction and parameters of motion are changed. On the example of a snooker game, motion of the ball is limited by other balls, cushion (table borders) and pockets.

The existing approaches for snooker game analysis, [2],[9],[6], aim at predicting the state of the balls in the current frame knowing its positions in previous frame(s). When the tracked data of the whole video is collected, they select a subsample for the tasks of summarization or event detection. In contrast, we propose that tracking every position of the target with known motion model is not influential in sense of video understanding. For this purpose, we predict the evolution of object's spatio-temporal changes according to its physical properties (motion functionality). This prediction is then taken in conjunction with the structure of the scene, in order to detect critical points of the trajectory and correct the parameters of original motion model.

The remaining of the paper is organized as follows. Next section introduces the idea of our approach. Section 3 is dedicated to the utilization of the proposed method for the application of tracking a snooker game. Each important issue is followed by the discussion of the state of the art approaches. The paper is concluded in Section 4.

## 2 Motion as a descriptor for tracking

One of the main components in the tracking pipeline is dedicated to modelling an object of interest, or shortly a target. There is no universal formal description of the key object properties that enable successful tracking in all possible cases. Generally, computer vision community is split into the adherents of the statistical and structural approach. The common feature for both approaches is that they do not work with the real world objects, but with the image representations and their properties.

In our view the model of the real object should not be limited to the properties of the image. Tracking aims at analyzing dynamic scenes, which in turn reveal the spatio-temporal changes of the objects. These dynamic changes are mostly not an unpredictable phenomena with independent random states in each moment. By the nature motion behaviour of an object has limitations that basically relate to motion functionality of the object and/or to the structure of the scene (see Table 1).

Functionality, in our opinion, represents a set of abilities that the object possesses. From motion perspective, functionality defines how the object could deform with time according to its physical nature. On one hand, it is defined by the structure of the object (e.g. rigid/non-rigid parts, degrees of freedom). On

**Tabelle 1.** Criteria impacting the motion behaviour

| **motion limitations** | | | |
|---|---|---|---|
| **functionality** | | **structure of the scene** | |
| structure of the object | motion model | space limits | obstacles |

the other hand, it depends on the local motion models of the constituent parts and the global motion model of the whole object.

Structure of the scene is a set of objects and conditions influencing the motion trajectory of the target. We consider obstacles and space limits to be of higher importance. Obstacle is an object of the scene which lies on the trajectory of the moving target. As a result the original motion model of the target is changed either after a collision or due to bypassing. Obstacle is not a permanent motionless part of a structure of the scene and may change its position. In contrast, space limits are constant, they constraint further motion of the target, such that bypassing is not feasible. Though collision and reflection scenarios are possible.

The awareness about the above phenomena provides advantages as opposed to other sources of guidance in the following current tracking problems:

- keep tracking in case of a partial or/and complete occlusion;
- prediction of the critical points of trajectory;
- lower the computational costs since processing only meaningful data.

**Multiple Facets of Motion Functionality** Prediction based on functionality makes a crucial impact and importance in a wide range of applications varying from computer graphics to robotics.

In robotics time is of critical importance since interacting with dynamically changing environment. Since 1980s the motion model of the flying ping pong ball is used for its trajectory prediction [3, 1], in order to configure the paddle for a successful ball return. Another example relates to a motion planner, where the incorrectness of the robot's navigation may lead to an injury of a human. Robot uses the predicted trajectories of the moving human, in order to prevent the collision risks [10].

Reconstructing facial surgery operation is a complex, radical (drastic) and irreversible procedure. The functionality of facial muscles enables accurate biomechanical facial soft tissue modelling and post-operative result simulation. Recent achievements in this direction relate to preoperative simulation of craniofacial surgery on bones with respective soft tissue alterations [13]. Other areas of application are human face visualization and mimics recognition [5], where the functionality provides robustness for the approach.

Approaches based on Kalman filter[4] are widely used in navigation systems and computer vision. This recursive physics-based method supports the estimation of the current position of the object taking into consideration previous states, measurements of the current state and a Gaussian nature of noise.

The distinct feature of the proposed approach is that the motion of the object is not taken apart and is observed in conjunction with the structure of the scene. It enables not only to predict the future spatio-temporal states of the object according to the given motion model, but also to be aware of the critical points where the trajectory can deviate from original.

## 3 Analysing a Snooker Game

The intention of this section is to show the applicability of the proposed approach to the real problem. For this purpose we selected the snooker game footage analysis, but the same methodology can be used in other domains for tracking objects with defined motion model and structure of the scene.

### 3.1 Motivation

Snooker is a variety of pool played with 21 balls of 6 distinct colors and a white ball (cue ball). The goal is to pot the color balls with a cue ball in a particular order and gain more points than the opponent. While watching or playing this game people are not tracking the positions of the moving balls as the time flows. On the contrary they try to predict future positions and pay attention to the prominent ones. In order to achieve planned trajectories, players use the effect of ball spin, the reflection from the cushion, or both. The correct model of the ball's movement should consider physical properties of a cue stick, baze, table and balls as well as several forces: rolling resistance, sliding friction, self-rotation[8]. Also from the point of winning the game, it is not only important to pot the ball. In case of disadvantageous position it makes sense to create a loosing situation for the opponent while his turn. On the whole it is preferable to hit as less balls as it is needed, in order to be able to predict the next state of the game. Thus, complicated shots involving several balls are not frequently used since they are not well predictable. In this manner there are mainly two balls which are predominantly needed to be tracked - cue ball and the hit ball. From this hypothesis it is obvious that tracking positions of all the balls in all the frames is not needed. Moreover, tracking the moving balls in all the frames is redundant when this move has a predictable trajectory. Using the abstract concept of visual attention together with the structure of the scene and functionality for predicting time and location of critical points of the trajectory becomes natural.

### 3.2 State of the art in snooker video analysis

Sport video analysis is widely represented in computer vision community: from semantic event detection and summarization to computer-assisting referee systems. A frequent engineering approach is to combine several existing methods for solving a particular task. Thus, the tools that seem to work on different problems have issues in common. For example, Denman et al. [2] introduced several approaches for video parsing, event detection and shot activity summarization in snooker footage analysis. This includes table shots detection rested on geometry and Hough Transformation, tracking a cue ball using color-based particle filter and detecting pots by histogram analysis of pocket regions. On the basis of this work with a modified ball tracking method Rea et al. [9] build a system for semantic event detection. For 3D reconstruction from snooker video [6] consider ball movements to be of more semantic importance as opposed to players and cue stick. In this manner they apply detection, classification and tracking of the balls. Tracking the objects of interest is a building block in these tools. According to its definition [11], searching for the object of interest is performed in subsequent frames. In contrast, the idea of our approach is to predict the evolution of object's temporal changes using its physical properties.

### 3.3 Description

The above mentioned human attention strategy (see Section 3.1) is modelled using motion functionality and structure of the scene. Prediction of future trajectories of the moving balls in conjunction with the space limits and positions of obstacles guide the selection of the prominent parts of the scene with corresponding time slots.

**Functionality** Having the set of the balls $B$=$\{b_1,\ b_2,\ldots,\ b_{21},\ b_{cue}\}$, the functionality is defined by a set $b_i \in B = \{S, V, a\}$, where $S$ is a vector of positions of the ball in consecutive video frames ($s_i = (x_i, y_i), s_i \in S, i = \overline{1, N}$), $V$ - a vector of velocity values for consecutive pairs in vector $S$, $a$ - acceleration of the ball. The motion of the ball is constrained to uniformly accelerated linear model.

*Remark 1* The velocity values $v_i \in V$ are computed as the first derivative of distance with respect to time, and acceleration $a$ as a second derivative. Due to this fact, the minimum number of positions for prediction is 3. The quality of the ordinary footage does not allow to precisely measure the velocity for obtaining a real value of acceleration. For that we consider Gaussian distribution of acceleration for getting the most likeable value of it. When the velocity is close to 0, it is assumed that the target does not move.

*Remark 2* The velocity of the ball decreases in every consecutive time slot. The exception of this rule occurs when during cushion collision the vector of ball-spin is collinear to the vector of ball-movement. As a result both the rebound angle and rebound velocity increase.

*Remark 3* The accurate model of the ball's movement should consider physical properties of a cue stick, baze, table and balls as well as several forces: rolling resistance, sliding friction, self-rotation[8]. Particularly, the rotational component of the moving ball makes an impact on the resultant trajectory such that the rebound angle from the cushion is deviating from perfect reflection, and the motion parameters change.

In order to obtain the consecutive positions $S$ of the balls, traditional tracking approaches are combined in the following way. The first movement before each shot corresponds either to the cue ball, or to the cue stick. Optical flow technique is combined with the blob detection algorithm to get the moving parts on the table. When the motion is detected, we check this region for the white ball using a circular Hough transform combined with color thresholding. The idea of color thresholding is to find such a circle in a region, which has the highest density of pixels exceeding value of 200 at each of RGB levels. Next a "snapshot" of this ball is taken and in the next frames apply template matching technique. Template matching is a procedure of finding a region in an image, that correlates the most with the given template. Due to the changes of the template caused by perspective projection and occlusions by the players or other balls, we perform Kalman filter, in order to obtain comparably robust track.

**Structure of the scene** According to the definition in Section 2, structure of the scene contains obstacles and space limits. In this application, obstacles are represented by the balls other than the cue (white) ball. After the hit they change the trajectory of the cue ball and move under their own functionality. The space limits are represented by the area of the table and the billiard-pockets. Under the rules, balls cannot cross the table borders. After cushion collision the ball's trajectory is changed due to the laws of reflection. Billiard-pocket is a part of the structure of the scene where the ball's trajectory ends.

*Obstacles* The most intriguing part of the snooker is to observe the change of trajectories caused by balls collision which may lead to a successful pot. At the beginning of each shot all the balls except the white are static. Obstacle-ball becomes of importance when belonging to the trajectory of the moving ball. The route of the moving ball is represented as a binary mask:

$$I(x, y) = \begin{cases} 1, (x, y) \in route \\ 0, otherwise \end{cases}$$

where I(x,y) - value of a pixel at position (x,y). Having a prediction of the motion vector and the distance until the ball stops, the vertices of the quadrangular *route* are calculated in the following way. Two vertices, $P_1$ and $P_2$, represent the intersection

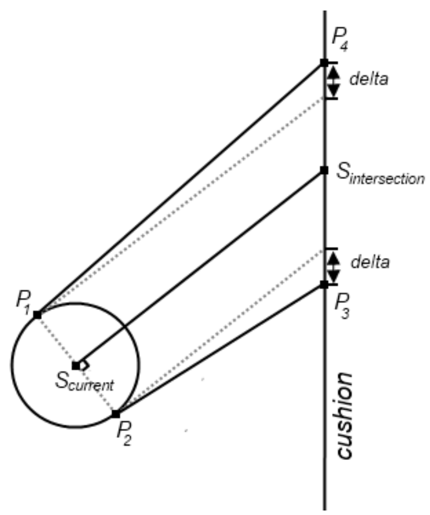


**Abb. 1.** Points estimation of quadrangular route in case of cushion collision

points between the circle of the ball and the line perpendicular to the motion vector passing through the center of the circle. The other two vertices, $P_3$ and $P_4$, are plotted on the line perpendicular to the motion vector passing through the end point of the trajectory $s_{predicted}$ at a distance ($radius + delta$) in both directions from the $s_{predicted}$, where *radius* - radius of the ball, *delta* - parameter that copes with small deviations of the ball's positions (see Figure 1). There are two cases for computing the values of $P_3$ and $P_4$ that should be distinguished. First, when the distance to the end point $s_{predicted}$ is smaller than the distance to the cushion. Second, when the distance to the end point $s_{predicted}$ is greater than the distance to the cushion. In the latter case, the intersection point between the trajectory and the cushion should be preliminary measured.

For the purpose of detecting, whether the obstacle-ball is inside the *route*, we multiply the binary masks of this ball and the *route*. In case the result is positive, we assume that there will be a collision, predict the time slot when it will happen, obtain template 'snapshot' of an obstacle-ball and start tracking it in the same way as with the cue (white) ball. If there exist several obstacle-balls on the *route*, those which will be hit first is taken into account.

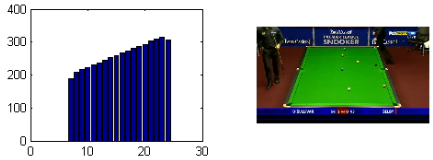


**Abb. 2.** Preliminary table detection based on RGB histogram approach

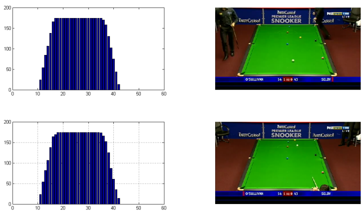

**Abb. 3.** Comparison of the first successful table view histogram with the current video frame histogram


**Abb. 4.** Finding the position of middle pockets as lying on the intersection of diagonals

<u>*Space limits*</u> In snooker broadcasts the effect of presence and involvement in a game is created via multiple camera views. In this paper we are particularly interested in a full-table view from the top (see Figure 4). The reason is that it has sufficient information for predicting a trajectory of the moving object. As opposed to other camera positions, it provides clear representation of the scene and is not dependant on 3D information for correct time and location estimations. Identification of such full-table view shots and parameters of the table is a vital step in obtaining the structure of the scene.

With the purpose to obtain the parameters of the full-table view frames, we initially accomplish a histogram approach in HSV color space. First, color thresholding and morphological closing are applied, in order to get a binary image of the green areas of the shot. For the resultant image 8-neighbor-connected components are found and the largest of them is assumed to be a candidate for a perspective view of the table. According to the perspective projection, the amount of green color increases with approaching to the bottom of the image. A candidate region is, finally, tested on satisfying this criteria.

When the first successful frame is detected, the corresponding candidate region is utilized to collect the information about the table – histogram, boundaries and pocket positions. The boundaries of the table are obtained by Hough transform as it finds the most prominent lines on the given binary image. After that the intersection points between the boundaries are assumed to be the corner table pockets. Two pockets in the center are estimated as the intersections between lateral boundaries of the table and a straight line parallel to the remaining boundaries through the intersection point of diagonals[2] (see Figure 4). For the upcoming video frames we manage a histogram comparison with the first frame. The above procedure is illustrated in Figures 2− 3. In case one of the pockets lies on the *route* of the moving ball, it is then further analysed for potting.

**Prediction** This part of the paper is dedicated to method of predicting the trajectory of the moving ball. It is assumed that the motion of the ball is limited to uniformly accelerated linear model. Having a track of a moving ball $S = \{s_0, s_1, .., s_k\} = \{(x_0, y_0), (x_1, y_1), .., (x_k, y_k)\}$, the relation between the parameters $x$ and $y$ is recovered using one-dimensional linear regression:

$$y_i = \alpha x_i + \beta + \epsilon_i, i = \overline{0, k}$$

$$(a, b) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=0}^{k} (y_i - \alpha x_i - \beta - \epsilon_i)^2$$

where $\alpha, \beta$ - motion model parameters (angular coefficient, absolute term); $\epsilon_i$-precision error; a,b - point estimates of $\alpha$ and $\beta$. We decided to restrict the motion to the linear model, though, the extension to non-linear model is possible.

The distance from the current position $s_{current} = (x_{current}, y_{current})$ with the velocity $v_{current}$ and acceleration $a$ until the stop of the ball is computed using the physical equation for uniformly accelerated motion:

$$s_{predicted} = \frac{v_{end}^2 - v_{current}^2}{2a} = \frac{0 - v_{current}^2}{2a}$$

Analysis of the video frames which correspond to the range $s_{current}$ and $s_{predicted}$ is eliminated.

**Experiments** Existing methods that aim at analysing snooker footage either do not provide the results for tracking, or only give a few details. In this situation a reasonable comparison of approaches is hardly achievable and, thus, we can only provide the summarization of our results.

This approach was tested on 12 minute snooker footage that is equal to 11480 video frames. Hereof 9598 frames(84%) contain full-table view, 2441 frames(21%) were neglected at motion detection step, 1758 frames(15%) were reduced using prediction component. In total about 36% of the video was not analysed. Overall processing took about 75 minutes. The tracker was lost in 161 frames(2%), which lead to incorrect route reducing of 62 frames(3%).

Additionally, we compared our approach with the results of tracking using a Kalman filter. It is worth reminding that Kalman provides a prediction/correction of the current target position. In contrast, our approach predicts/corrects the future important positions of the target which are collisions with the table borders or other balls. In general case, when the trajectory is closed to a straight line, the proposed approach enables faster analysis by neglecting at once in average 1-3 seconds of video (20-60 video frames) with an average deviation of 10 pixels (diameter of the ball is 8-12 pixels). The advantage of performing sequential tracking with Kalman filter can be shown on the example of curvilinear trajectory (Figure 5). In Figure 6 it is shown that Kalman corrects the position step by step and follows the real path. As opposed to it, our system predicts the linear future position and loses the robustness of a track.

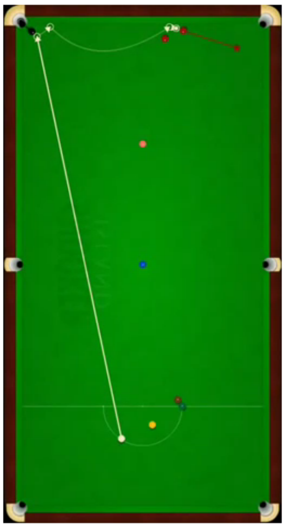

**Abb. 5.** The outline of the shot with curvilinear component

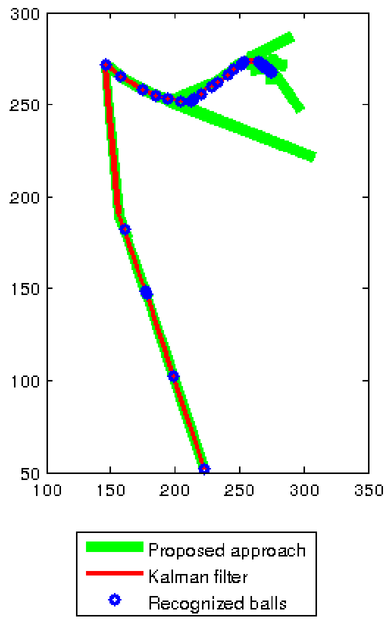


**Abb. 6.** The results of processing the motion of the ball with curvilinear component

## 4 Conclusion

This paper presented a framework for snooker video analysis. The future positions of the moving balls are predicted using physically-based linear motion model with respect to the structure of the scene. Motion model is characterized by notions of velocity, acceleration and previous states of the tracked object. Structure of the scene represents the obstacles and space limits that impact the trajectory and motion parameters of the target. In terms of snooker application they are cushion, pockets and balls other than the target. For the future work we plan to research the rotational component of the ball-motion. This feature makes a valuable impact on motion model, as well as on the reflection angle while hitting the cushion or other balls.

## References

1. Russell L Anderson. *A robot ping-pong player: experiment in real-time intelligent control.* MIT press, 1988.
2. Hugh Denman, Niall Rea, and Anil Kokaram. Content-based analysis for video from snooker broadcasts. *Computer Vision and Image Understanding*, 92(2):176–195, 2003.
3. Yanlong Huang, De Xu, Min Tan, and Hu Su. Trajectory prediction of spinning ball for ping-pong player robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2011*, pages 3434–3439. IEEE, 2011.
4. Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.
5. Dongwoon Lee, Michael Glueck, Azam Khan, Eugene Fiume, and Ken Jackson. A survey of modeling and simulation of skeletal muscle. *ACM Transactions on Graphics*, 28(4):1–13, 2010.
6. Philip A Legg, Matthew L Parry, David HS Chung, Richard M Jiang, Adrian Morris, Iwan W Griffiths, D Marshall, and Min Chen. Intelligent filtering by semantic importance for single-view 3d reconstruction from snooker video. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2385–2388. IEEE, 2011.
7. Emilio Maggio and Andrea Cavallaro. *Video tracking: theory and practice.* John Wiley & Sons, Chichester, United Kingdom, 2011.
8. Michael McCloskey. Naive theories of motion. *Mental models*, pages 299–324, 1983.
9. Niall Rea, Rozenn Dahyot, and A Kokaram. Semantic event detection in sports through motion understanding. In *Image and Video Retrieval*, pages 88–97. Springer, 2004.
10. Emrah Akin Sisbot, Luis Felipe Marin-Urias, Rachid Alami, and Thierry Simeon. A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883, 2007.
11. Richard Szeliski. *Computer vision: algorithms and applications.* Springer, 2011.
12. Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.
13. Martin Weiser, Stefan Zachow, and Peter Deuflhard. Craniofacial surgery planning based on virtual patient models. *it-Information Technology*, 52(5):258–263, 2010.
14. Jeremy M Wolfe. Guided search 4.0. *Integrated models of cognitive systems*, pages 99–119, 2007.