MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

3D Trajectory Reconstruction of the Soccer Ball for Single Static Camera Systems

Jürgen Metzler and Frank Pagel
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
Fraunhoferstrasse. 1, 76131 Karlsruhe, Germany
{juergen.metzler,frank.pagel}@iosb.fraunhofer.de

Abstract

The interest on acquiring player and ball data during soccer games is increasing in several domains such as media or training. Consequently, tracking systems are becoming widely used for live data gathering. However, due to costs, stadium infrastructure, media rights etc. there is a trend for stand-alone mobile low-cost soccer tracking systems. The drawback of such systems is that generally only low-resolution images of the players are available. Thus, the problem of detection and tracking the soccer ball is also strongly exacerbated. This paper presents a novel approach for 3D reconstruction of the ball trajectory from monocular low-resolution soccer image sequences. The ball detection is done on accumulated motion-segmented binary images by extracting shape-validated ball candidates from these so-called motion history images. Then, robust ball tracklets are extracted from the motion history images and used to reconstruct the ball 3D trajectory based on physical characteristics and calibration information. The approach is tested on a Bundesliga data set: the tracklet extraction and the ball trajectory reconstruction are evaluated.

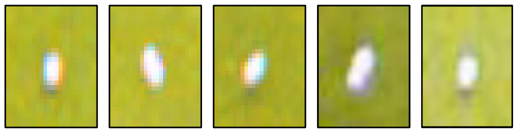

Figure 1. Variety of the ball's appearance. All samples are extracted from one image sequence.

1 Introduction

The increasing professionalization of soccer is accompanied by a growing media attention as well as game analysis and professional training. Especially the automation of live analysis of soccer games is interesting for several domains such as media or training. However, the automation presupposes a robust acquisition of player and ball data that still relies heavily on the interaction of operators (so-called scouts) in current systems. Live acquisition of quantitative motion data such as distance covered by players or ball possession can only be done by sophisticated automation. Our overall tracking system provides this kind of quantitative data for supporting a scout and for the automated acquisition of the relevant data [1]. It automatically detects, classifies and tracks the ball, the 22 soccer players, the referee and the two linesmen in one image sequence of double Full HD resolution.

The main contribution of this work is the acquisition of the ball trajectory based on ball detections. Detection of the ball in image sequences generally is a difficult task as appearance of the ball varies from image to image. For instance, the high accelerations occurring at the ball may cause motion blur so that the appearance of the ball is then more of an ellipse than a circle (see some examples in Fig. 1). Also, the color of the ball may be vary from image to image because of changes of the illumination conditions or it has the same color as the lines of the pitch which exacerbates the ball detection task. Another challenge is the image resolution of the ball which is usually very small so that also confusions with body parts may occur. Depending on the camera perspective, the ball is in front of a complex image background such as the audience which exacerbates its detection as well. Besides difficulties that arise from the appearance of the ball by itself, the detection of the ball is very challenging in situations where it is occluded by the players.

The approach presented in this contribution is applicable for static camera systems, even for low-resolution cameras. So it can be used in huge tracking systems consisting of several cameras usually fix installed in stadiums as well as for low-cost camera systems that generally consist of 1-3 cameras capturing the entire pitch.

There are a couple of publications for detecting and tracking the soccer ball [4]. However, many of them tracking the ball in broadcast soccer videos. One approach processing images from static cameras and using accumalated foreground segmented images as well, is presented in [3]. After the foreground is segmenetd, connected foreground regions being too large for a ball are deleted and accumulated with the last n segmented images. The accumulated image is then used to find the initial ball position and for density estimation of ball hypotheses for tracking the ball by using partical filtering.

For 3D reconstruction of the ball trajectory, the calibration information describing the relative position and orientation of the camera relative to the soccer field plays an important role [6, 2]. While multi-camera solutions just need the ball tracks from different camera views to be able to reconstruct the 3D trajectory [7, 9], single camera applications mostly assume physical models like parabolas for the 3D reconstruction [6, 2]. Especially the points where the ball touches the ground are critical information for estimating the models [7, 6, 2]. Unlike tracking applications for volley ball games [2] or golfing [9], in the soccer scenario we have to deal with lots of occlusions (where the tracked ball merges with players) and incomplete data in terms of missing ball trajectory points. In our approach, we do not need to specifically determine the end points of the parabola *a priori*. Furthermore, we determine

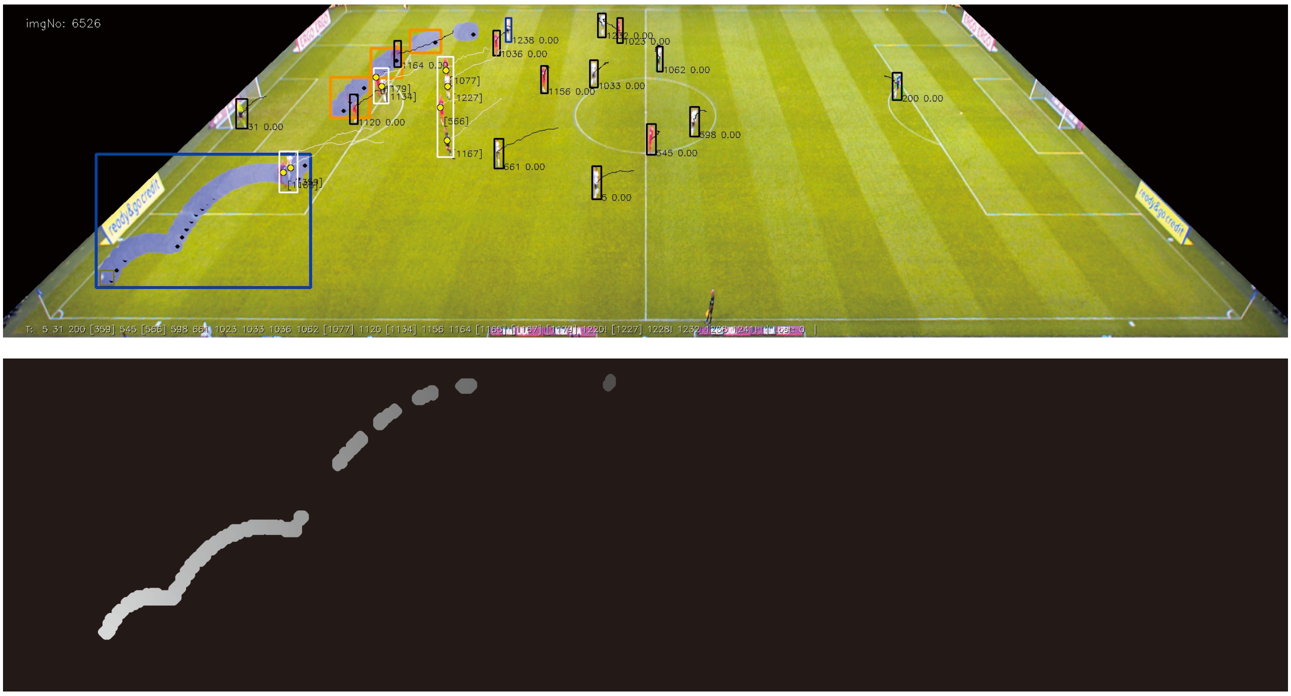

Figure 2. Sample snapshot of the processed input image sequence including detected ball tracklets (top) and one for the according motion history image (bottom).

those points implicitely and hence segment the ball trajectory by estimating the physical motion model.

In this contribution, a novel approach for detecting the ball and reconstructing its trajectory in monocular low-resolution soccer image sequences is presented. Ball candidates are first detected in motion-segmented binary images by size and validated by analyzing their shape. Candidates that are not confirmed due to their shape are removed from the binary images. Then, the binary images with the remaining and validated ball candidates are accumulated to a motion history image (MHI), where latest images are additionally weighted higher.. Finally, the reconstruction of the ball trajectory is made by a physically motivated model-based approach based on the linked robust ball tracklets extracted from the MHI.

The contribution is structured as follows: in Section 2, the module for the ball detection is proposed. Then, the calibratin of the cameras is described in Section 3 and the reconstruction of the ball trajectory in Section 4, before the results are shown in Section 5.

2 Ball detection

A reliable soccer ball detection simplifies the ball trajectory reconstruction obviously. Therefore, the false alarm rate of the detections should be kept low. However, because the ball is the main object of interest there are a lot of occlusions and the changing appearance of the ball makes this task difficult. In order to achieve a high detection rate at a low false alarm rate, we follow a detection approach that has been widely established: we divide the detection task in two steps. First, ball candidates are extracted followed by an add-on verification of them.

After image acquisition, the soccer ball has to be detected first. Due to the real-time constraint, a feature-based detection with e.g. a sliding windows approach cannot be used. Instead, as the images are captured by a static camera, a detection of ball candidates is performed by foreground/background segmentation first [5]. Temporal static background like the pitch and marking lines are segmented as background, whereas moving objects generate changing appearance and therefore foreground segmentation. During the ball candidate extraction, all foreground regions are extracted and checked for its size using calibration information of the cameras. Foreground regions that are no candidates for the soccer ball due to their size are removed.

At the second step, external contours of the remaining regions are extracted as a sequence of points and analyzed . First, an ellipse is fitted to the sequence and the mean squared error between every sequence point and the ellipse is calculated. Ball candidates with a high mean squared error are removed. Using ellipses, deformations of the ball due to, for instance, offsets in camera synchornization can be handled. The remaining candidates are kept as verified foreground regions in the foreground/background segmented image. Then, a dilatation is applied and the last n binary images are accumulated to a MHI of verified ball candidates, before the tracklets are finally extracted from it (see Fig. 2).

3 Calibration

In order to be able to determine the position of the ball and the players on the pitch as well as the height of ball tracks, we calibrate the cameras intrinsically and extrinsically [8]. For the extrinsic calibration the soccer field itself is used, as the size of

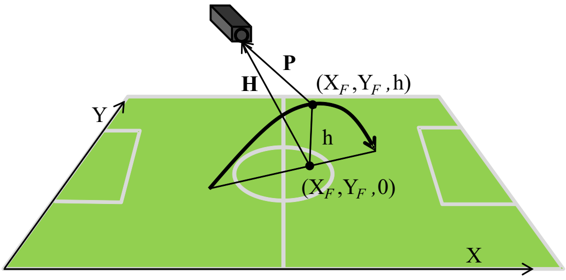

Figure 3. The calibration parameters can be used to determine the correspondences between image points and points on the ground of the playing field, and based on those ground points, the 3D height of an image point.

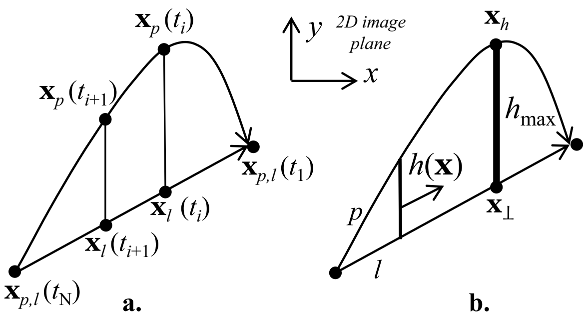

Figure 4. 2D ball tracklets in the image plane. **a**. Scheme of the relationships to determine the temporal consistency for both, the parabola and the line model. **b**. The peak of the parabola is detected by vertically sweeping along the parabola. The maximum distance between line l and the parabola p indicates the peak.

the pitch and lengths of lines are usually well known. Hence, as a result we get the camera matrix $\mathbf{P}_{3\times 4} = (\mathbf{p}_1^T, \mathbf{p}_2^T, \mathbf{p}_3^T)^T = (p_{ij})$ which describes the projection of a 3D point onto the image plane in homogeneous coordinates: $\tilde{\mathbf{x}} = \mathbf{P}\tilde{\mathbf{X}}$ with $\tilde{\mathbf{x}} = (\mathbf{x}^T, 1)^T = (x, y, 1)^T$ and $\tilde{\mathbf{X}} = (\mathbf{X}^T, 1)^T = (X, Y, Z, 1)^T$. Furthermore, we can determine the homography $\mathbf{H}_{3\times 3}$ that maps pixel coordinates in the image onto 2D coordinates of the playing field: $(X_F, Y_F, 1)^T = \mathbf{H}\tilde{\mathbf{x}}$ (see Fig. 3).

4 3D trajectory reconstruction

We start with a list of tracked ball positions and try to decide which points of the recent trajectory belong either to a pass (ball remains on the ground) or a cross (ball is in the air) or nothing of both. In the latter case we assume a parabola as the underlying model of the ballistic trajectory of the ball.

The trajectory of length N can be incomplete due to missing detections or occlusions. So, we iteratively check track segments of length N_k starting with the latest trajectory point and increasing the segment by sequentially adding points from the past. Hereby, we evaluate the segment regarding the line model and the parabola model. In a first step, we estimate the model parameters $(a_l, b_l)^T$ for a line and $(a_p, b_p, c_p)^T$ for a parabola with

$$\begin{array}{rcccl} y & = & a_l x + b_l & =: & \rho_l(x) \text{ and} \\ y & = & a_p x^2 + b_p x + c_p & =: & \rho_p(x). \end{array} \tag{1}$$

Using the 2D points from the current trajectory segment we can formulate an overdetermined linear equation system and solve for the model parameters. Afterwards, we check for the temporal consistency of the segment. This means, that according to the calculated model the order of the points must be sequential and the temporal appearances of ball detections $\mathbf{x}$ at time t must be consistent. Therefore, we assume linear motion of the ball along the estimated trajectory model, which leads to the following condition for temporal consistency of two consecutive trajectory points $\mathbf{x}(t_i), \mathbf{x}(t_{i+1})$ in a segment of length N_k for a line and a parabola respecively (see Fig. 4. a):

$$\begin{array}{lcl} \left\| \Delta t \frac{\|\mathbf{x}(t_1)-\mathbf{x}(t_{N_k})\|_2}{\Delta T} - \Delta_{S_l} \right\|_2 & - \tau(\Delta t)\Delta_{S_l} & \overset{!}{<} 0 \\ \left\| \Delta t \frac{\langle \mathbf{x}(t_1),\mathbf{x}(t_{N_k})\rangle_p}{\Delta T} - \Delta_{S_p} \right\|_2 & - \tau(\Delta t)\Delta_{S_p} & \overset{!}{<} 0 \end{array} \tag{2}$$

with

$$\begin{array}{rcl} \Delta t & = & t_i - t_{i+1}, \\ \Delta T & = & t_1 - t_{N_k}, \\ \Delta_{S_p} & = & \langle \mathbf{x}(t_i), \mathbf{x}(t_{i+1})\rangle_p, \\ \Delta_{S_l} & = & \|\mathbf{x}(t_i) - \mathbf{x}(t_{i+1})\|_2 \qquad \text{and} \\ \tau(t) & = & \alpha t^2. \end{array} \tag{3}$$

The neglect of air resistance will be payed attention by introducing a tolerance function $\tau(t)$ with a tolerance factor α, which allows deviations from the model depending on the temporal duration of the track segment. We choose a square function to be able to also cover rolling balls that continously reduce speed and hence slowly moving towards a rigid position. The lengths of the parabola segments are approximated by a piecewise segment accumulation:

$$\begin{array}{c} \langle \mathbf{x}(t_1), \mathbf{x}(t_2)\rangle_p = \\ \Sigma_{i=0}^{L} \left\| \begin{pmatrix} x_1 + i\Delta x \\ \rho_p(x_1 + i\Delta x) \end{pmatrix} - \begin{pmatrix} x_1 + (i+1)\Delta x \\ \rho_p(x_1 + (i+1)\Delta x) \end{pmatrix} \right\|_2 \end{array} \tag{4}$$

with $\Delta x = \frac{|x(t_1) - x(t_2)|}{L}$ and L sample points.

This iterative model estimation and evaluation will be repeated until the longest possible segment is found that fits a model. If a segment fulfills both conditions for a line and a parabola, we choose the line model. If only the parabola model is valid, we continue estimating the parabola model in the following time steps until a new model is found, which means that we have detected a complete parabola trajectory segment. This is important as for the height estimation of the 3D parabola we need the complete parabola with its two end points and not just a piece in between. In cases of too strong occlusions (e.g. after a corner kick) it may happen that no model can be applied.

Once we detected a complete parabola segment we continue estimating the height of it. Therefore, we use the calibration information. Assuming that we know the two end points of the parabola in the image, we can determine the line l that goes through these two points.

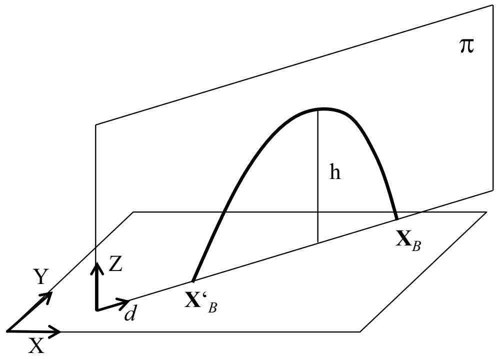

Figure 5. 3D parabola model of the ball trajectory based on the estimated height h of the ball trajectory and the end points.

By vertically sweeping through the parabola segment, we determine the parabola peak $\mathbf{x}_h$ to be the point where the vertical distance between l and the parabola is maximal (see Fig. 4. b). The respective point on l is $\mathbf{x}_\perp$. As we know the point $\mathbf{x}_\perp$ in the image plane and the calibration parameters, we can calculate the corresponding point on the soccer field $\tilde{\mathbf{X}}_F = \mathbf{H}\tilde{\mathbf{x}}_\perp$. Now, given a ball with its coordinates over ground $\mathbf{X}_F = (X_F, Y_F)^T$, we are looking for the height h that best describe the projected point $\mathbf{x}_h$. Based on the projection equation $\tilde{\mathbf{x}}_h = \mathbf{P}(X_F, Y_F, h, 1)^T$ from Section 3, we can solve for h with

$$h = \frac{y_h \mathbf{p}_3^T \tilde{\mathbf{Y}} - \mathbf{p}_2^T \tilde{\mathbf{Y}}}{p_{23} - y_h p_{33}}. \tag{5}$$

with $\tilde{\mathbf{Y}} := (X_F, Y_F, 0, 1)^T$. Finally, given the height h and the bordering end points $\mathbf{X}_B, \mathbf{X}'_B$ in the ground plane, the 3D parabola of the ball trajectory in a plane π orthogonal to the ground plane (determined by the the end points and the normal $(0,0,1)^T$) can be approximated by $\rho_\pi(d) = d^2 + a_\pi d + h$ with $a_\pi = -\frac{D^2-4h}{2D}$ and $D = \|\mathbf{X}_B - \mathbf{X}'_B\|$ (see Fig. 5).

5 Results

We tested the proposed approach on a data set of a Bundesliga match. It consists of an image sequence with about 140.000 images of double Full HD resolution (see Fig. 2 for an example). There are 1428 tracklets to detect, in situations where the ball is not occluded or not merged with a player. The approach detected 1343 tracklets and missed 85. There was no false alarm, i.e. all detcted ball tracklets were correctly deteced as such.

The accuracy of the estimated height of the ball strongly depends on the accuracy of the estimated ball positions where the ball touches the ground. Furthermore, due to the low resolution of the video material, we are not able to automatically determine if a ball moves back in the air before it touches the ground (e.g. a header). As we did not have ground truth for the actual ball height, we used the hand-labeled ball trajectories to verify the correctness of the determined ball segments. We could observe a very satisfying segmentation in the majority of the scenes. However, we could also observe false positives, which means we over-segmented a ball trajectory, in the following three situations: First, when the ball is highly passed almost parallel to the center line from far to close or vice versa. Second, when the assumed parabolic motion model does not meet the true ball motion, e.g. due to air friction or ball spin. And third, when the tracklets were too fractured, e.g. when the distance between the players and hence the length of the passes is very short.

6 Conclusion

In this paper, an approach for 3D trajectory reconstruction of the ball for monocular low-resolution soccer image sequences has been presented. We could yield a reliable extraction of ball tracklets and estimation of the position of the ball on the pitch as well as its height estimation in 3D space. Furthermore, by using a motion history of the ball and a physical motion model, we could implicitely segment the motion history of the ball into straight passes on the ground and high crosses. With the exception of the delay in the output of the ball coordinates, which depends on the length of the motion history, the proposed approach is real-time capable.

References

[1] C. Herrmann, D. Manger, and J. Metzler: "Feature-based Localization Refinement of Players in Soccer Using Plausibility Maps," *in Proc. of the IPCV Conf., WORLDCOMP*, Vol. 2, 2011.

[2] H.-T. Chen and C.-L. Chou and W.-J. Tsai and S.-Y. Lee: "3D ball trajectory reconstruction from single-camera sports video for free viewpoint virtual replay," *in Proc. of the IEEE VCIP Conf.*, pp. 1–4, 2011.

[3] K. Choi and Y. Seo: "Probabilistic Tracking of the Soccer Ball," *LNCS Statistical Methods in Video Processing, ECCV Workshop SMVP*, Vol. 3247, pp. 50–60, 2004.

[4] T. D'Orazio and M. Leo: "A review of vision-based systems for soccer video analysis," *Pattern Recognition*, Vol. 43, No. 8, pp. 2911–2926, 2010.

[5] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis: "Real-time foregroundbackground segmentation using codebook model," *Real-Time Imaging In Special Issue on Video Object Processing*, Vol. 11, No. 3, pp. 172–185, 2005.

[6] T. Kim and Y. Seo and K.-S. Hong: "Physics-based 3D Position Analysis of a Soccer Ball from Monocular Image Sequences," *in Proc. of the IEEE ICCV*, pp. 721–726, 1998.

[7] J. Ren, J. Orwell, G. A. Jones and M. Xu: "A General Framework For 3D Soccer Ball Estimation And Tracking," *in Proc. IEEE Int. Conf. on Image Processing*, 2004.

[8] Z. Zhang: "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1330–1334, 2000.

[9] T. Zupančič and A. Jaklič: "Automatic Golf Ball Trajectory Reconstruction and Visualization," *in Proc. of the 4th Int. Conf. on Computer Vision/Computer Graphics Collaboration Techniques*, pp. 150–160, 2009.