

Automatic Camera Pose Recognition in Planar View Scenarios

Luis Alvarez, Luis Gomez, Pedro Henriquez, and Luis Mazorra

CTIM (Centro de Tecnologías de la Imagen),
Universidad de Las Palmas de Gran Canaria, Spain
{lalvarez,lgomez,phenriquez,lmazorra}@ctim.es
<http://www.ctim.es>

Abstract. The main goal of this paper is to recognize automatically camera pose from a single image of a planar view scenario. We apply this technique to sport event scenarios using as information the white lines/circles dividing the different parts of the sport court. Using these court primitives we define a loss function that we minimize to obtain the best perspective transformation (homography) matching the actual sport court with its projection in the image. From such homography we recover the camera pose (position and orientation in the 3D space). We present numerical experiments in simulated and real sport scenarios.

Keywords: camera calibration, sport scenarios, lens distortion.

1 Introduction

Camera calibration is an important issue in computer vision. In particular, the broadcasting of sport events increasingly introduces the processing of video sequences for a wide variety of purposes, such as mosaicing, change of the view point or insertion of virtual objects. Some of these tasks require a highly precise calibration of the cameras. In this paper we address the problem of automatic camera pose recognition in the context of planar view scenarios as sport events where the court the sport is played in a planar surface with known dimensions and with a number of lines or circles (usually in white) dividing the different parts of the court. This is a quite common situation in sport scenarios, for instance, tennis, basketball or soccer courts satisfied this assumption. The court primitives we use to recover the camera pose are the white lines/circles dividing the different parts of the sport court.

As we will show below, when we take a photo of the planar sport court, the position of the court primitives in the image is given by a perspective transformation (an homography) of their actual position. In figure 1 we illustrate this perspective transformation. We will show in this paper that if a minimum number of court primitives are visible in the scene then we can recover the homography which transforms the image scene court to the actual reference court. Then from this homography (and assuming that some internal camera parameters are known) we can recover the camera pose, that is the position and orientation of

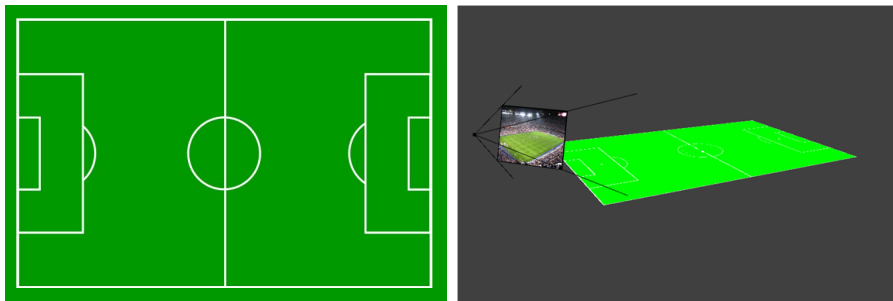


Fig. 1. On the left, illustration of actual Soccer Court. On the right Camera Pose.

the camera. So the method we propose in this paper to automatically recognize and recover camera pose can be divided in the following steps:

1. Court primitives extraction (white lines/circles) in the image.
2. To define and minimize a loss function, using the extracted court primitives, which allow us to evaluate the quality of the matching between actual and image court for an specific perspective transformation (homography)
3. To recover the camera pose from the estimated homography.

The organization of the paper is as follows : in section 2, we present some related work. Section 3 is devoted to briefly explain the camera model we use. In section 4 we study the problem of court primitives extraction. In section 5 we introduce the loss function we use to perform camera pose recognition and we design methods to estimate, automatically, the homography which minimizes the loss function. In section 6 we study how to recover camera pose from the homography. In section 7 we present some numerical experiments on simulated and real scenarios. Finally, in section 8 we present some conclusions.

2 Related Works

Most line extraction methods applied to sport events commonly use a segmentation with dominant colour detection using HSV space [1] or Gaussian mixture models [2]. But to be able to extract lines when dealing with interlaced images, HD definition images, or scenarios with significant contrast variations between the background and the lines, we use the method described in [3]. In the following step, we need to estimate the lens distortion model to obtain an accurate evaluation of the quality of the matching between actual and image court. To deal with lens distortion we use the approach detailed in [5,6]. Finally, to calculate the extrinsic parameters of the camera (rotation and focus) some methods based on soccer lines/circles have been proposed as [4],and using court models as in [7,8]. Also, there are another methods which use the vanishing points of the lines as [9].

3 Camera Model

The camera model we use in this work is called pinhole model, where we include a radial lens distortion model. Using this camera model, a 3D point (X, Y, Z) is projected in the image in a 2D point (x, y) in projective coordinates given by:

$$\begin{pmatrix} d_x(x, y) \\ d_y(x, y) \\ 1 \end{pmatrix} = sAR \begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (1)$$

Where R is the rotation matrix and $C = (c_x, c_y, c_z)$ is the focus. They are called extrinsic parameters. A is the intrinsic parameters 3×3 matrix and $(d_x(x, y), d_y(x, y))$ is the lens distortion correction model.

4 Court Primitives Extraction

We obtain image court primitive points using standard algorithms to extract centre line points on the white image court primitives (see for instance [3]). So our starting point is the set P given by :

$$P = \{p_i = (x_i, y_i) \in \text{Image court primitives}\} \quad (2)$$

We apply the Hough transform to extract lines from the set P . In figures 7 of the numerical experiments section we illustrate the collection of extracted lines in different images. We observe that some of the lines corresponds to tangents to circles projection in the image. Although, initially, these tangents can be considered as faults in Hough transformation line estimation, in fact, the information of such tangents can be use to extract the associated ellipse equation. To build the ellipse equation from the tangents we use the nice and classic work [11] (published in 1885). In this work, different methods of building ellipse equation are shown using different kind of information. In particular, using this work approach, we can easily deduce that using 2 tangents of the ellipse with their correspondence contact points and a third line passing by the ellipse centre we can extract ellipse equation. So using this technique we can extract ellipse from the image using just line information. Of course this is much faster and robust than to look for ellipses in the image using standard techniques based on using just primitive point information.

5 Loss Function: Definition and Minimisation

Let us note by C the collection of actual court primitives. Given an homography H (a 3×3 matrix) and a 2D point p we denote by $H(p)$ the perspective transformation induced by H on point p . We define the loss function $L(H)$ as :

$$L(H) = \sum_{p_i \in P} \text{distance}(H(p_i), C)^2 \quad (3)$$

Find the minima of the above loss function is a quite difficult problem. The method we propose is based on building homography candidates using the collection of primitive lines we have extracted in the image. We observe (see the numerical experiments) that the number of lines we manage is quite small, so the number of potential configuration we have to manage is also small. We separate our analysis in 2 cases according to the scene configuration:

Case 1: There are at least four visible primitives lines in the image. In this case we build homography candidates by putting in correspondence 4 visible lines in the scene with 4 lines in the actual court. We observe that for each pair of lines put in correspondence between the image court and the actual court, we obtain the relation:

$$\begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}^T \begin{pmatrix} a'_i \\ b'_i \\ c'_i \end{pmatrix} = S \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} \quad (4)$$

From the above relation we deduce 2 linear equations in the coefficients of H

$$(a'_i h_1 + b'_i h_4 + c'_i h_7) b_i = (a'_i h_2 + b'_i h_5 + c'_i h_8) a_i \quad (5)$$

$$(a'_i h_1 + b'_i h_4 + c'_i h_7) c_i = (a'_i h_3 + b'_i h_6 + c'_i h_9) a_i \quad (6)$$

Collecting the above linear equations for the 4 line correspondence we obtain the equation system $Bh = 0$ where B is a 8×9 matrix, The system solution is obtained by minimising $\|Bh\|^2$, and we can deduce that the homography H is given by the minimum eigenvector of $B^T T B$ matrix.

Case 2: The centre circle and the half-way line are visible in the image. Additionally, it will be supposed that it is also visible one touchline or the court centre spot. This case is more complex to analyze and its solution is based on work done by Luis Alvarez and Vicent Caselles in [10], where they show a method to recover the homography from this scenario (we refer to [10] for more details).

We observe that we do not have "a priori" information about the scenario configuration we deal with. So we check both configurations (case 1 and case 2) and for each configuration we build different homographies by considering potential combination of lines in the image and primitives in the actual court. Finally, we keep the homography with the loss function lowest value.

6 Camera Pose Recognition

We fix a 3D coordinate system where the actual court is included in the plane $z = 0$. To recover the camera pose we need to manage the intrinsic parameter matrix A and the homography H given by :

$$A = \begin{pmatrix} f & 0 & x_c \\ 0 & f \cdot r & y_c \\ 0 & 0 & 1 \end{pmatrix} \quad H = sAR \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & -c_z \end{pmatrix} \quad (7)$$

We assume that intrinsic camera parameters are known, except the focal distance f which varies according to zoom parameter setting. To recover the focal distance from H we observe than from the above equation we obtain:

$$H^T A^{-T} A^{-1} H = s^2 \begin{pmatrix} 1 & 0 & -c_z \\ 0 & 1 & -c_y \\ -c_z & -c_y & (c_y)^2 + 2(c_z)^2 \end{pmatrix} \quad (8)$$

Replacing A with its value we obtain :

$$H^T \begin{pmatrix} 1 & 0 & -x_c \\ 0 & \frac{1}{r^2} & -\frac{1}{r^2}y_c \\ -x_c & -\frac{1}{r^2}y_c & \frac{1}{r^2}y_c^2 + x_c^2 + f^2 \end{pmatrix} H = s \begin{pmatrix} 1 & 0 & -c_z \\ 0 & 1 & -c_y \\ -c_z & -c_y & (c_y)^2 + 2(c_z)^2 \end{pmatrix} \quad (9)$$

With $b = (b_1, b_2, b_3, b_4) = (\frac{1}{r^2}, x_c, -\frac{1}{r^2}y_c, -\frac{1}{r^2}y_c^2 + x_c^2 + f^2)$. From the previous equation we obtain two equations in the b coefficients:

$$\begin{aligned} h_{21}(b_1h_{21} + b_3h_{31}) + h_{11}(b_2h_{31} + h_{11}) + h_{31}(b_2h_{11} + b_3h_{21} + b_4h_{31}) - \\ h_{22}(b_1h_{22} + b_3h_{32}) + h_{12}(b_2h_{32} + h_{12}) + h_{32}(b_2h_{12} + b_3h_{22} + b_4h_{32}) = 0 \\ h_{22}(b_1h_{21} + b_3h_{31}) + h_{12}(b_2h_{31} + h_{11}) + h_{32}(b_2h_{11} + b_3h_{21} + b_4h_{31}) = 0 \end{aligned} \quad (10)$$

As b_1, b_2 and b_3 are known, the only unknown is b_4 , therefore, from only one camera it is possible to calculate b_4 and subsequently the focal distance f .

Next, to calculate the extrinsic parameters (rotation and focus) from the homography and the intrinsic parameters, we use the following relationships: The homography H , that projects the reference field on the image field, is shown in expression 7. From that equation we can obtain the following expression:

$$R = sA^{-1}H \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & -c_z \end{pmatrix} = sA^{-1}H \begin{pmatrix} 1 & 0 & -\frac{c_x}{c_z} \\ 0 & 1 & -\frac{c_y}{c_z} \\ 0 & 0 & -\frac{1}{c_z} \end{pmatrix} \quad (11)$$

Equating the first two columns of the matrixes in expression 11 we obtain the rotation matrix. Equating the third column we obtain the focus.

We will assume that the internal parameters of the camera such as the pixel aspect ratio and the projection of the camera focus in the camera projection plane are known.

7 Numerical Experiments and Results

We have tested our method on different images using both, scale models (1440 x 809 frames) and real scenes from soccer matches (1920 x 1080). Moreover, we have tested the method in the 2 different cases, when at least 4 lines are visible and when the centre circle, the half-way line and a touchline are visible. In the table 1 we can see the results obtained when we applied the method to an image of each case. The second column contains the number of points which



Fig. 2. Scale model and real images: Images on top are the detected lines. Bottom images are illustrations of the estimated pose of the camera in each top image.

have been labelled as a primitive point with the morphological thick line center detection [3]. In the third column we can see the number of lines extracted after applying the Hough transform. Finally, we show the loss function average, measured in metres, and we can see that the results are very accurate. Indeed, in the worst case we obtain only a 2 centimetres error, being usual an error of some milimetres.

On the other hand, we have used the estimated camera pose to build a 3D representation of each camera. In the figure 2 some examples are shown. In that images we can observe image pairs, where one of them (top ones) is the real image with the detected lines painted in different colours, and the other one (bottom ones) is a 3D illustration with the estimated camera pose represented by a image miniature and 4 lines determining the view. We can see with those images that the camera position and orientation have been properly calculated.

Table 1. Quantitative results for images in Fig. 2. we show the number of court primitives points and lines used and the loss function point average measured in metres

Images from Fig. 2	Points	Lines	Loss function average
Area (scaled court model image)	4146	9	5.5921e-004
Centre (scaled court model image)	2705	9	8.0006e-004
Corner (real court image)	5986	12	9.8619e-003
Area (real court image)	4977	11	8.9835e-003
Centre (real court image)	3174	10	3.7095e-002
General view (real court image)	3880	10	2.9171e-002

8 Conclusions

In this paper we study the problem of camera pose estimation in scenarios where there are usually a small number of visible primitives which can be considered to perform the estimation. We have shown that if a minimum number of court primitives are visible in the scene then we can recover the homography which transforms the image scene court to the actual reference court. We have proposed a method based on building homography candidates using the collection of primitive lines we have extracted in the image to find the homography which minimizes the loss function. Then from this homography we can recover the camera pose, that is the position and orientation of the camera. We present some experiments using HD images of soccer matches in both, scale models and real scenarios. The numerical results we present are precise and we show some examples with 3D illustration of the estimated camera poses.

Acknowledgement. This research has partially been supported by the MICINN project reference MTM2010-17615 (Ministerio de Ciencia e Innovación. Spain). We acknowledge MEDIAPRODUCCION S.L. for providing us with the real HD video we use in the numerical experiments.

References

1. Ekin, A., Tekalp, A.M., Mehrotra, R.: Automatic Soccer Video Analysis and Summarization. *IEEE Transactions on Image Processing* 12(7) (July 2003)
2. Liu, Y., Huang, Q., Ye, Q., Gao, W.: A new method to calculate the camera focusing area and player position on play field in soccer video. In: *Visual Communications and Image Processing* (2005)
3. Alemán-Flores, M., Alvarez, L., Henríquez, P., Mazorra, L.: Morphological Thick Line Center Detection. In: Campilho, A., Kamel, M. (eds.) *ICIAR 2010. LNCS*, vol. 6111, pp. 71–80. Springer, Heidelberg (2010)
4. Li, Q., Luo, Y.: Automatic Camera Calibration for Images of Soccer Match. *Proceedings of World Academy of Science, Engineering and Technology* 1, 170–173 (2005)
5. Alvarez, L., Gomez, L., Sendra, J.R.: An algebraic approach to lens distortion by line rectification. *Journal of Mathematical Imaging and Vision* 35, 36–50 (2009)
6. Alvarez, L., Gomez, L., Sendra, J.R.: Accurate Depth Dependent Lens Distortion Models: An Application to Planar View Scenarios. *Journal of Mathematical Imaging and Vision* 39, 75–85 (2011)
7. Farin, D., Krabbe, S., de With, P.H.N., Effelsberg, W.: Robust camera calibration for sport videos using court models. *Storage and Retrieval Methods and Applications for Multimedia* 5307, 80–91 (2004)
8. Jiang, B., Songyang, L., Liang, B.: Automatic line mark recognition and its application in camera calibration in soccer video. In: *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6 (2011)
9. Babae-Kashany, V., Reza Pourreza, H.: Camera Pose Estimation in Soccer Scenes Based on Vanishing Points. In: *9th IEEE International Symposium on Haptic Audio-Visual Environments and Games*, pp. 157–162 (2010)
10. Alvarez, L., Caselles, V.: Calibration method for a TV and video camera. *European Patent 09380137.1* (issued, February 2011)
11. Eagles, T.H.: *Constructive geometry of plane curves*. Macmillan and Co. (1885)