

Article

Camera calibration in sport event scenarios

Aleman-Flores, M, Alvarez, L, Gomez, L, Henriquez Castellano, Pedro and Mazorra, L

Available at <http://clock.uclan.ac.uk/18033/>

Aleman-Flores, M, Alvarez, L, Gomez, L, Henriquez Castellano, Pedro and Mazorra, L (2014) Camera calibration in sport event scenarios. Pattern Recognition, 47 (1). pp. 89-95. ISSN 0031-3203

It is advisable to refer to the publisher's version if you intend to cite from the work.

<http://dx.doi.org/10.1016/j.patcog.2013.05.011>

For more information about UCLan's research in this area go to <http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Camera Calibration in Sport Event Scenarios

M. Alemán-Flores^a, L. Alvarez^a, L. Gomez^b, P. Henriquez^{a,*}, L. Mazorra^a

^a*CTIM: Centro de I+D de Tecnologías de la Imagen. Departamento de Informática y Sistemas. Universidad de Las Palmas de Gran Canaria. Campus de Tafira, 35017, Las Palmas. Spain.*

^b*CTIM: Centro de I+D de Tecnologías de la Imagen. Departamento de Ingeniería Electrónica y Automática. Universidad de Las Palmas de Gran Canaria. Campus de Tafira, 35017, Las Palmas. Spain.*

Abstract

The main goal of this paper is the design of a novel and robust methodology for calibrating cameras from a single image in sport scenarios, such as a soccer field, or a basketball or tennis court. In these sport scenarios, the only references we use to calibrate the camera are the lines and circles delimiting the different regions. The first problem we address is the extraction of image primitives, including the challenging problems of shaded regions and lens distortion. From these primitives, we automatically recognise the location of the sport court in the scene by estimating the homography which matches the actual court with its projection onto the image. This is achieved even when only a few primitives are available. Finally, from this homography, we recover the camera calibration parameters. In particular, we estimate the focal length as well as the position and orientation in the 3D space. We present some experiments on models and real courts which illustrate the accuracy of the proposed methodology.

Keywords: Camera calibration, Sport scenarios, Camera pose.

*Corresponding author: Tel. +34 928 458708

Email addresses: `maleman@ctim.es` (M. Alemán-Flores), `lalvarez@ctim.es` (L. Alvarez), `lgomez@ctim.es` (L. Gomez), `phenriquez@ctim.es` (P. Henriquez), `lmazorra@ctim.es` (L. Mazorra)

1. Introduction

During the broadcast of sport events, video sequences are usually processed 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, that is, to know the *camera pose*: position and orientation of the camera and its intrinsic parameters (focal length, pixel size and the principal point). Moreover, the estimation of the lens distortion is likely to be needed for the case of highly precise camera calibration. In what follows, we will assume that all intrinsic parameters of the camera are known except for the focal length, which can be modified by the camera operator by changing the camera zoom. Such assumption comes from the fact that cameras at sports events are usually mounted on fixed tripods and are previously calibrated by means of standard off-line calibration methods to extract the intrinsic parameters. During the broadcast of the sport event, the cameras remain in the initial fixed position and it is even usual that they remain at the same position for different matches played at the same court.

In sport scenarios the main primitives we can use to calibrate the camera are the white lines and circles dividing the sport court. We assume that the court is a planar surface with known dimensions and a certain number of lines or circles (usually in white colour) dividing the different parts of the court. This is a quite common situation in sport scenarios, such as tennis, basketball or soccer.

As we will show below, when we capture a picture of a planar scenario, the positions of the primitives in the image are given by a homography. In Fig. 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 plane into the actual reference plane. 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 the camera.

In the application of our methods, we focus on soccer images. These are one of the most challenging types, since they present natural light conditions (which are therefore quite variable), shaded regions, and heterogeneous background.

The main contribution of the paper is the introduction of a novel and fully automatic technique that allows estimating, from a single image, the camera calibration parameters in complex scenarios where we deal with a

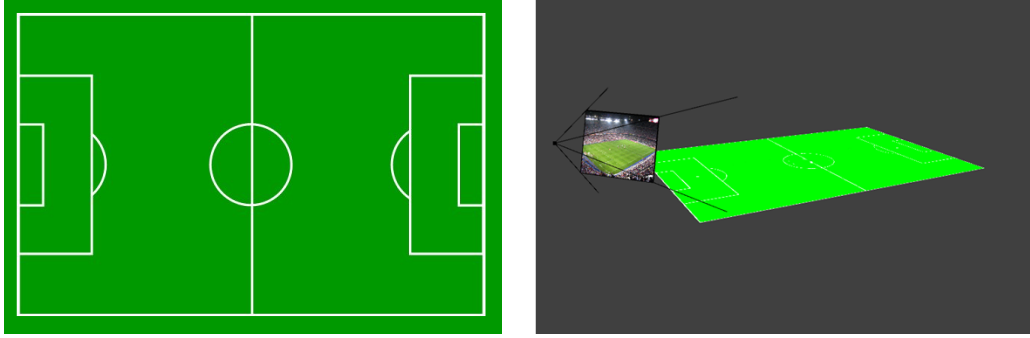


Figure 1: Reference soccer field (left), and camera pose reconstruction showing the acquired image and the corresponding position and orientation of the camera with respect to the field (right).

few visible primitives, poor light conditions, shaded areas or lens distortion. The proposed technique can be divided into the following steps:

1. Primitive extraction (white lines/circles) in the image. We also consider the challenging problem of images showing large shaded areas. In sport events broadcasted under natural light conditions (*daylight*), we can deal with large areas where the primitives are shaded, which makes the extraction of the primitives a more difficult task.
2. Sport court recognition. We automatically estimate the homography which transforms the actual sport court into its projection onto the image. To do that, we propose some novel strategies to match the primitives extracted from the image and the actual primitives in the reference plane.
3. Camera calibration. We recover the camera calibration parameters from the estimated homography. In Fig. 1, we illustrate the result of the calibration procedure.

The organisation of the paper is as follows: in Section 2, we present some related works. Section 3 is devoted to primitive extraction. In Section 4, we introduce the camera model which has been used. Section 5 presents the loss function which measures the accuracy of a homography. Section 6 is devoted to camera calibration. In Section 7, we present some experiments on simulated and real scenarios. Finally, in Section 8, we present some conclusions.

2. Related works

Camera pose determination is a major issue in Computer Vision. The problem was formulated by Fischler and Bolles [1] for the case of a camera with constant focal length, that is, no zooming was accounted, and the first attempts to solve it used 2D or 3D point correspondences [2, 3]. Other works consider more features into the recognition phase, such as corners or man-made objects [4], [5]. These methods exploit some knowledge within the scene (e.g. lines or corners) or even use standard calibration patterns (*chessboard*) and require multiple views. Then, a nonlinear optimisation problem is stated and solved by gradient-like optimisation methods to get a satisfactory local minimum. Radial distortion or the possibility of changing the camera zoom are not usually taken into account in order to simplify the problem, or are included in a post-processing stage. In this work, to deal with lens distortion we use the algebraic approach detailed in [6, 7], where the authors use a least square approximation of those edges which should be a projection of 3D lines as a measure of distortion. Afterwards, they calculate the sum of the squares of the distances from the points to the lines to obtain the distortion error.

A method for self-calibration of rotating and zooming cameras is described in [8]. Other works present some methods for multi-camera calibration, such as [9, 10]. Based on those techniques, some optimisations for the calibration of PTZ (Pan-Tilt-Zoom) cameras have been introduced [11], and there are also some methods that do not require predefined patterns for tracking, such as [12].

There has been a great deal of recent works in the pattern recognition research community tackling the specific case of the camera pose problem for planar scenarios. Video tracking for soccer matches, automatic analysis of soccer video sequences and summarisation of complete soccer matches require a precise camera calibration and camera pose recognition [13], [14], [15], [16], [17].

For the case of planar scenarios, there is a significant reduction of the complexity of the problem due to the fact that the camera is mounted on a fixed tripod and the correspondence to be established is also reduced to a 2D problem [18]. One of the first published works researching the particular problem of camera calibration under the geometrical 2D restriction was [19]. In this manuscript, the author shows that the camera pose for a moving projective camera can be estimated from a minimum number of 5 frames of a planar view under the constraint that the projections of orthogonal di-

rection vectors (points at infinity) in the plane must be orthogonal in the calibrated camera frame of each image. In [20], an automatic camera calibration for short sequences of video soccer matches taken in the penalty area near one of the goals is introduced. The authors apply the well-known Tsai's camera calibration method [21] and a planar transformation (*homography*) is computed from some primitives extracted from the images to estimate the intrinsic and extrinsic camera parameters, including the zoom factor. In [22], by using images taken simultaneously by two PTZ cameras which are positioned at different locations and observe the same dynamic event taking place on a planar surface, a planar homography is calculated for each time instant using the image correspondences between both images and, from that homography, the camera pose is obtained.

Most line extraction methods applied to sport events commonly use a segmentation with dominant colour detection using HSV space [13] or Gaussian mixture models [23]. In this paper we apply the method described in [24] for extracting lines when dealing with interlaced images, HD definition images, or scenarios with significant contrast variations between the background and the lines.

Finally, to calculate the extrinsic parameters of the camera (rotation and focus) some methods have been proposed using lines/circles [25] or court models [26, 27]. There are also some other methods which use the vanishing points of the lines, such as [28].

3. Primitive extraction

In order to calibrate the camera, we need to extract the primitives which will help us identify its parameters. These are the white lines and circles which characterise a soccer pitch. To this aim, we will make use of the following disk morphological operators:

Given an image $I : \Omega \rightarrow \mathfrak{R}$ and a disk D_s centred in $(0, 0)$ and with radius s , we define:

$$\begin{aligned} \text{Disk dilation :} \quad & I \oplus D_s(x) = \sup_{y \in x + D_s} I(y) \\ \text{Disk erosion :} \quad & I \ominus D_s(x) = \inf_{y \in x + D_s} I(y) \\ \text{Disk opening :} \quad & I \circ D_s(x) = (I \ominus D_s) \oplus D_s(x) \\ \text{Disk closing :} \quad & I \bullet D_s(x) = (I \oplus D_s) \ominus D_s(x), \end{aligned}$$

where $x \in \Omega$, \sup and \inf are the supremum and infimum, respectively. When applying a disk opening operator $I \circ D_s$, any white line such that its thickness is lower than the disk diameter ($2s$) is removed from the image. Therefore,

if we take the disk diameter larger than the maximum white line thickness presented in the RGB image, the subtraction between the original image and the opening, $I - I \ominus D_s$, will provide a segmentation of such white lines. We set a threshold on the difference $I - I \ominus D_s$ and consider the following region A:

$$A = \left\{ x : \begin{cases} (R(x) - R \circ D_s(x)) > t_R \\ (G(x) - G \circ D_s(x)) > t_G \\ (B(x) - B \circ D_s(x)) > t_B \end{cases} \right\}, \quad (1)$$

where t_R, t_G, t_B are thresholds for the different channels. As the thick line area is very small with respect to the line background area, the parameters t_R, t_G and t_B are chosen in terms of a percentage $0 < p < 1$ with respect to the histogram of the corresponding image channel. For instance, t_R is chosen to satisfy:

$$p = \frac{|x \in C : (R(x) - R \circ D_s(x)) > t_R|}{|C|}, \quad (2)$$

where $|\cdot|$ represents the cardinal (size) of the set. In the experiments we chose $p = 0.02$.

Region A (Eq.1) will contain the image white lines on the field. However, it also contains all those narrow regions which are considerably brighter than their neighbourhoods, even if they are located outside the pitch. Therefore, we must restrict the process to the soccer field, so that the rest of the image is not taken into account. The main difference between the soccer field and the rest of the image is found in the hue channel, which remains within a certain interval all across the field. This feature is useful to extract the region of interest and, therefore, instead of working on the RGB channels, we turn to the HSV representation and consider the hue channel. In this channel, we select an interval $[t_{H_1}, t_{H_2}]$ which comprises the range of typical values for the grass. Let $(H_s(x), S_s(x), V_s(x))$ be the HSV channels of the image $(R \circ D_s, G \circ D_s, B \circ D_s)$. Then, the field area F can be expressed as:

$$F = \{x : t_{H_1} \leq H_s(x) \leq t_{H_2}\}, \quad (3)$$

where t_{H_1} and t_{H_2} are chosen analysing the peak of the histogram of H_s channel using a standard histogram segmentation technique (see [29] for details).

Since H_s is computed after the opening process, the lines are included in F . As a result of the operation $A \cap F$, we obtain the final set \mathcal{B} , which

contains those line points which correspond to the image primitives located in region of interest.

These lines have a certain width, which can be quite variable, but we are interested in obtaining one-pixel wide lines in order to use them for the calibration. Morphological skeleton will be used to find out the thick lines centres. Given a set X , the morphological skeleton is given by:

$$S = \cup_{s>0} (\cap_{\mu>0} (X \ominus D_s \setminus (X \ominus D_s) \circ D_\mu)), \quad (4)$$

where D_s is a disk of radius s centred at $(0, 0)$. The skeleton represents, for a given shape X , the centres of the maximal disks included in X .

In the case of discrete lattices, the morphological skeleton can be stated in the following way: Let D_n be the disk of radius n centred in $(0, 0)$. Then, the centre points of the line of width n can be obtained as the set:

$$S_n = (\mathcal{B} \ominus D_n) \setminus ((\mathcal{B} \ominus D_n) \circ D_1), \quad (5)$$

where \mathcal{B} is the detected thick line region. Therefore, the skeleton computation also provides, automatically, the line width.

3.1. Primitive extraction in shaded regions

In scenarios with large shaded regions, the method presented above does not work properly because a single threshold configuration for t_R , t_G and t_B cannot be set for both, the light and shaded areas. These situations can automatically be identified by considering the distribution of $V_s(x)$, i.e. the histogram of the value or brightness, within the previously determined region of interest. Usually, the hue of the background area does not change significantly from the light to the shaded areas, but the value channel $V_s(x)$ in the HSV space does. If we are dealing with just one region, the histogram of the values $h(V_s)$ has a single peak, whereas it presents two peaks if light and shaded regions are analysed together. Using a standard histogram segmentation technique (see [29] for details) we can automatically identify the number of significant peaks in the profile of $h(V_s)$. Once we have separated the shaded and lighted regions, we apply the same procedure proposed above to each region and then combine the results into a common line region \mathcal{B} for the whole image.

3.2. Court primitive extraction: the particular case of the ellipse.

From the above methodology, the points belonging to the court primitives are extracted. Next, using a standard Hough transform we detect lines in the image. In Figs. 4, 5 and 6, we illustrate the collection of lines extracted in different images. As observed, some of the lines which have been extracted are not exactly lines, but tangents to circles which are projected as ellipses on the image. Although, initially, these tangents could be considered as faults related to the lines extracted by the Hough transform, they provide us with some useful information about the associated ellipse equations. These equations can be obtained from the tangents using the method described in [30] (*published in 1885*). In that work, several methods for building ellipse equations are shown using different kinds of information. In particular, we can easily obtain the equation using two tangents of the ellipse with the corresponding contact points and a third line passing through the ellipse centre. Therefore, only line information is needed and it is certainly much faster and more robust than looking for ellipses using only the point information available.

4. Camera model

The camera model we use in this work is the pinhole model. To deal with the radial distortion, a polynomial lens distortion model has been included. Using this camera model, a 3D point (X, Y, Z) is projected onto the image in a 2D point (x, y) in projective coordinates given by:

$$\begin{pmatrix} d_x(x, y) \\ d_y(x, y) \\ 1 \end{pmatrix} = sAR \begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad (6)$$

where the rotation matrix, R , and the camera focus, $C = (c_x, c_y, c_z)^T$, correspond to the camera extrinsic parameters. The 3×3 matrix A contains the camera intrinsic parameters, $(d_x(x, y), d_y(x, y))$ is the lens distortion correction model, and s is an arbitrary scale factor. Note that, although the pinhole model is quite simple, it is also widely applied in computer vision. Additionally, the inclusion of the radial distortion model allows dealing with real camera lens aberrations.

4.1. Lens distortion model

The lens distortion model used is briefly presented in this subsection. The basic standard model for barrel and pincushion distortion correction (see for instance [31], [32] or [33]) is the simple radial distortion model given by the following expression:

$$\begin{pmatrix} \hat{x} - x_c \\ \hat{y} - y_c \end{pmatrix} = L(r) \begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix}, \quad (7)$$

where (x, y) is the original (distorted) point, (\hat{x}, \hat{y}) is the corrected (undistorted) point, (x_c, y_c) is the centre of the camera distortion model (the centre of the image), $r = \sqrt{(x - x_c)^2 + (y - y_c)^2}$ and $L(r)$ is the function which defines the shape of the distortion model. $L(r)$ can be approximated by a Taylor expansion, as follows:

$$L(r) = k_0 + k_1 r^2 + k_2 r^4 + \dots, \quad (8)$$

where the set $\mathbf{k} = (k_0, k_1, \dots, k_{N_k})^T$ consists of the distortion parameters which must be estimated from image measurements usually applying non-linear optimisation techniques.

The complexity of the above radial distortion model is given by the number of terms of the Taylor expansion we use to approximate $L(r)$. The one-parameter model, as [34] reports, achieves an accuracy of about 0.1 pixels in image space using lenses exhibiting large distortion. The two-parameter model is the usual approach in some related works (see for instance [35], or [6]) because it is simple, accurate and can be estimated using just image information (known straight lines in the scene provide enough information for simple cases). In particular, Alvarez, Gomez and Sendra [6] proposed an algebraic method to compute lens distortion models by correcting the line distortion induced by the lens distortion. This algebraic model is suitable for correcting small and strong radial distortion and is also highly efficient in terms of computational cost. For all the images shown in the experiments and results section, this algebraic model has been used to estimate and to correct the radial distortion. An on-line demo of the implementation of this algebraic method can be found in [36].

5. Loss function: definition and minimisation

As we may be dealing with many configurations and candidate homographies, a certain measure of the accuracy of each candidate is needed in order

to select the most appropriate one. Let us denote by P the collection of primitives in the actual court. Given a homography H (a 3×3 matrix) and a 2D point p , we denote by $H(p)$ the perspective transformation induced by H on the point p . We define the loss function $L(H)$ as :

$$L(H) = \frac{1}{|S|} \sum_{p \in S} \text{distance}(H(p), P)^2, \quad (9)$$

where *distance* refers to the Euclidean distance and S is the set which contains all points of the detected primitives. Finding the minima of this loss function is a difficult problem. The method we propose is based on building suitable homography candidates using the collection of primitive lines and ellipses extracted from the image. We observe (see Section 7) that the number of lines to deal with is small, and so is the number of potential configurations to consider. We separate our analysis in 2 cases according to the scene configuration:

Case 1: There are at least four visible lines in the image. In this case we build homography candidates by putting in correspondence (*matching*) 4 visible lines in the scene with 4 lines in the actual court. For each matched pair of lines the following relation holds:

$$\begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}^T \begin{pmatrix} a'_i \\ b'_i \\ c'_i \end{pmatrix} = s \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix}. \quad (10)$$

From the above relation, 2 linear equations in the coefficients of H are obtained:

$$(a'_i h_1 + b'_i h_4 + c'_i h_7) b_i = (a'_i h_2 + b'_i h_5 + c'_i h_8) a_i, \quad (11)$$

$$(a'_i h_1 + b'_i h_4 + c'_i h_7) c_i = (a'_i h_3 + b'_i h_6 + c'_i h_9) a_i. \quad (12)$$

Assembling this linear equations for the 4 matched lines, we obtain the linear system $Bh = 0$, where B is a 8×9 matrix. The solution of this system, that is, the homography H , is obtained by minimising $\|Bh\|^2$ and selecting the minimum eigenvector of the $B^T B$ matrix.

Case 2: The centre circle and the half-way line are visible in the image. Additionally, we assume that either a touchline or the centre spot is also visible. This case is more complex to analyse than the previous one and its

solution is based on the work done by Luis Alvarez and Vicent Caselles in [37], where a method to recover the homography from this scenario is shown.

As there is no “a priori” information about the initial scenario configuration, we propose the following procedure: both configurations (previous case 1 and case 2) are checked and solved, and then, the homography with the minimum loss function value is selected.

6. Camera pose recognition

In order to recover the camera pose, we first set the 3D coordinate system with the actual court lying on the plane $z = 0$. We assume that all intrinsic camera parameters (matrix A) are known, except for the focal distance f which varies according to the zoom parameter setting. Additionally, the matrix A and the estimated homography H satisfy the relation:

$$A = \begin{pmatrix} f & 0 & x_c \\ 0 & f \cdot r & y_c \\ 0 & 0 & 1 \end{pmatrix}, \quad H = sAR \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & -c_z \end{pmatrix}. \quad (13)$$

From the above equations, we can recover the focal distance, taking into account that:

$$H^T A^{-T} A^{-1} H = s^2 \begin{pmatrix} 1 & 0 & -c_z \\ 0 & 1 & -c_y \\ -c_z & -c_y & (c_y)^2 + 2(c_z)^2 \end{pmatrix}, \quad (14)$$

and replacing A with its value, we obtain:

$$H^T \begin{pmatrix} 1 & 0 & b_2 \\ 0 & b_1 & b_3 \\ b_2 & b_3 & b_4 \end{pmatrix} H = s \begin{pmatrix} 1 & 0 & -c_z \\ 0 & 1 & -c_y \\ -c_z & -c_y & (c_y)^2 + 2(c_z)^2 \end{pmatrix}, \quad (15)$$

where $b = (b_1, b_2, b_3, b_4) = (\frac{1}{r^2}, -x_c, -\frac{1}{r^2}y_c, \frac{1}{r^2}y_c^2 + x_c^2 + f^2)$. Then, from the previous expression, we obtain two equations in the b coefficients and, as b_1, b_2 and b_3 are known, the only unknown is b_4 . Therefore, from only one image it is possible to calculate b_4 and subsequently the focal distance f , that is, all the intrinsic parameters (matrix A) can be determined.

Next, to compute the extrinsic parameters (camera rotation and camera focus) from the homography and the intrinsic parameters, we take into account that, from Eq. (13), we obtain :

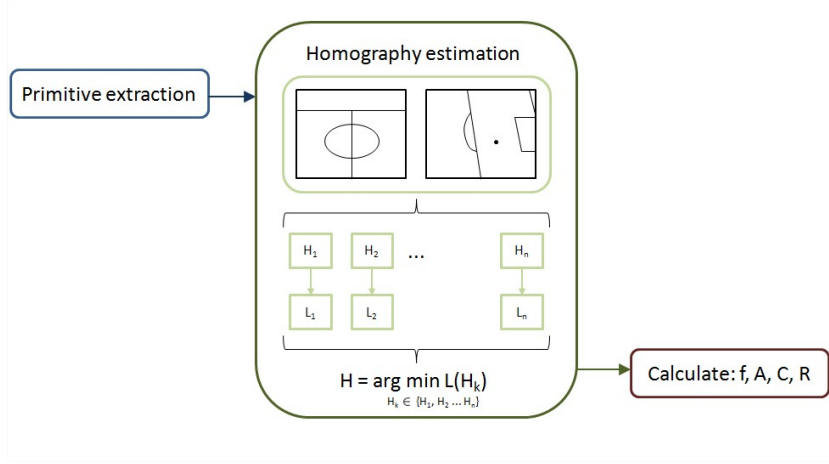


Figure 2: Calibration process stages.

$$R = sA^{-1}H \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & -c_z \end{pmatrix} = sA^{-1}H \begin{pmatrix} 1 & 0 & -\frac{c_x}{c_z} \\ 0 & 1 & -\frac{c_y}{c_z} \\ 0 & 0 & -\frac{1}{c_z} \end{pmatrix}. \quad (16)$$

Equating the first two columns of the matrices in expression (16) we obtain the rotation matrix (note that R is an orthonormal matrix). Next, by equating the third columns, we obtain the focus.

Therefore, it has been shown that by applying this procedure, a single camera can be calibrated in a straightforward manner and, hence, the camera pose can fully be determined. The whole calibration process is summarised in Fig. 2. As it can be seen in this figure, once the primitives are extracted (*primitive extraction* task), first, the two possible configurations (corresponding to the cases described in Section 5) are considered, and the set $\mathbf{H} = \{H_1, H_2, \dots, H_n\}$ of candidate homographies matching the actual and the reference court is obtained. From each element of this set, a loss function value (Eq. (9)) is obtained and the homography with the minimum loss function value is selected. From that homography and proceeding as it has been explained above, the camera pose is determined.

Images	Shot type	Points	Lines	Loss function
Scale model	Area	4146	9	5.5921e-004
	Centre	2705	9	8.0006e-004
Real court	General view	3880	10	2.9171e-002
	Centre	3174	10	3.7095e-002
Shaded regions	Area	641	9	1.5515e-002
	Area (from corner)	601	9	2.8088e-002

Table 1: Quantitative results for images in Figs. 4, 5 and 6: Number points and lines of the court primitives used for the calibration and loss function, measured in metres.

7. Experiments and results

In this section, we show some quantitative and qualitative results of our approach. To this aim, the proposed method has been tested on different images using both, scale simulated models (1440×809 images) and real scenes from soccer matches (1920×1080 images). Not only the normal conditions have been considered, but also some scenes with shaded regions. We have tested the method in two different cases, depending on the primitives available for the calibration. The first one corresponds to those scenes in which at least four lines are visible. In the second case, the centre circle, the half-way line and a touchline are visible.

The quantitative results are presented in Table 1, which includes several situations, views and conditions in both, the scale court and a real scenario. This table contains the number of points which have been labelled as a primitive point with the morphological thick line centre detection [24], as well as the number of lines extracted by using the Hough transform. Finally, the value of the loss function (see Eq. (9)), which is measured in metres, proves the accuracy of the results. An example of the obtained numerical results for one particular image, is shown in Fig. 3.

The estimated camera calibration has been used to build a 3D representation of each camera pose with respect to the field. By using OpenGL (Open Graphics Library), we create a virtual soccer field and locate a virtual camera in the corresponding position and orientation, in such a way that the view acquired by this camera matches the real image.

In Fig. 4, we can see the results for two scale model scenes. From the input image (top), we obtain the camera pose and represent it in a 3D

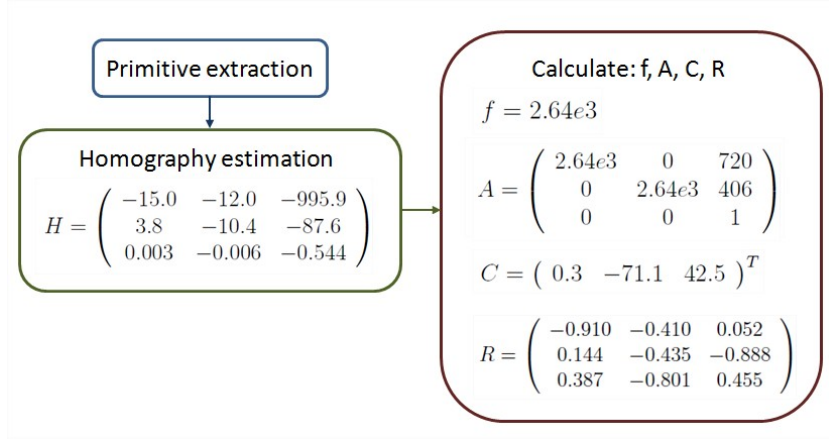


Figure 3: Results obtained for the calibration process (Top left image on Fig. 4)

illustration of the field (bottom). On the original image, we have drawn the lines which have been detected and used for the calibration. In Fig. 5, the same procedure is applied to real scenes of soccer matches. Finally, in Fig. 6, we show the robustness of the method under difficult conditions, such as those cases in which a part of the field is shaded by the stands. As observed, the camera pose extracted by means of this method is coherent with the view in the original images. With these qualitative examples, we prove that the camera position and orientation can be properly calculated.

8. Conclusions

In this paper we study the problem of camera calibration estimation in sport event scenarios where we have to deal with some difficult problems, such as a reduced number of visible primitives, poor light conditions, shaded areas or lens distortion. We propose a novel and fully automatic method which addresses all these issues and allows recovering the camera calibration parameters from a single image. This method includes a robust technique to estimate primitives in the image. We use lines as main basic primitives and we propose recovering the equation of the ellipses from their tangent lines (this approach strongly simplifies the problem: it is enough to detect lines). We also propose a homography estimation technique, using a loss function, that allows to automatically recover the perspective transformation from the

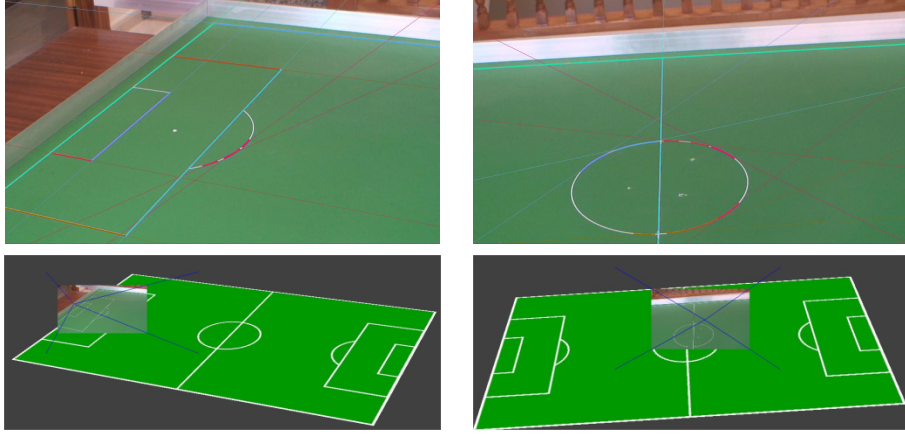


Figure 4: Scale model: Area (left) and centre (right). Lines detected on the field (top), and reconstruction of the estimated camera pose, showing its position and orientation with respect to the field (bottom).

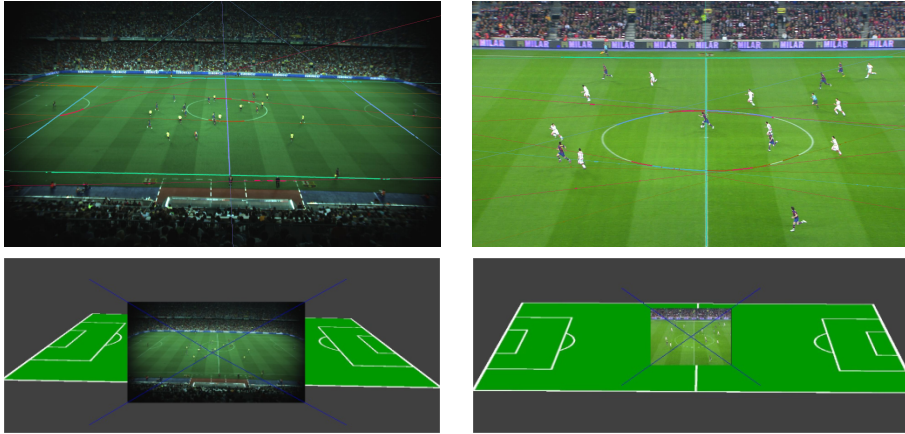


Figure 5: Real court: General view (left) and centre (right). Lines detected on the field (top), and reconstruction of the estimated camera pose, showing its position and orientation with respect to the field (bottom).

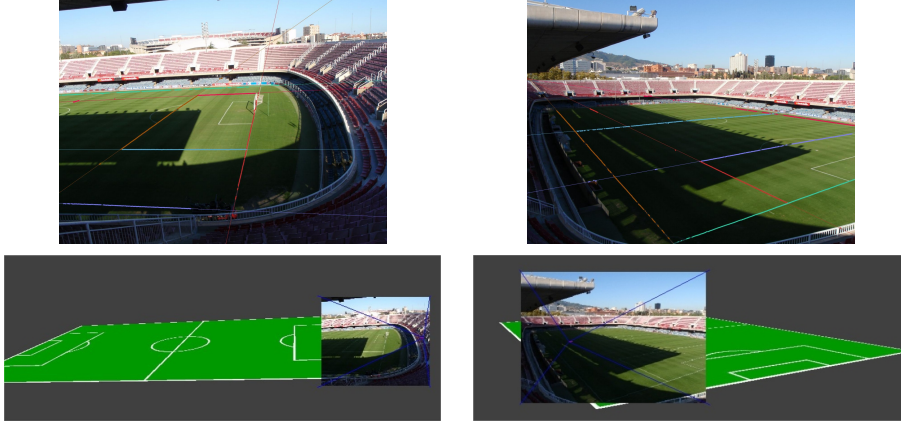


Figure 6: Real court with shaded regions: Area (left) and area from corner (right). Lines detected on the field (top), and reconstruction of the estimated camera pose, showing its position and orientation with respect to the field (bottom).

actual reference soccer pitch onto the image. Finally from the estimated homography we recover the camera focal length and pose.

To show the validity of the proposed method we have presented some experiments using HD images of soccer matches in both, scale models and real scenarios. The experiments have been performed in a variety of situations like central views (where only the central circle, the half-way line and a touch line are visible), panoramic views, penalty area views and images presenting large shaded regions. The experimental results we present are very promising and we have shown some 3D representations of the estimated camera poses to illustrate the results.

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 experiments.

References

- [1] M.A. Fischler, R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM 24 (6) (1981) 381-395.

- [2] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11) (2000) 1330-1334.
- [3] A. Ansar, K. Daniilidis. Linear pose estimation from points or lines. *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (5) (2003) 578-589.
- [4] X. Zhang, Y. Liu, T.S. Huang. Determining 3D structure and motion of man-made objects from image corners. In: *Proc. of SSIAI02 Santa Fe New Mexico, USA*, (2002) 26-30.
- [5] F. Shi, X. Zhang, Y. Liu. A new method of camera pose estimation using 2D3D corner correspondence. *Pattern Recognition Letters* 25 (2004) 1155-1163.
- [6] L. Alvarez, L. Gomez, J.R. Sendra: An algebraic approach to lens distortion by line rectification. *Journal of Mathematical Imaging and Vision* 35 (2009) 36-50.
- [7] L. Alvarez, L. Gomez, J.R. Sendra: Accurate depth dependent lens distortion models: An Application to Planar View Scenarios. *Journal of Mathematical Imaging and Vision* 39 (2011) 75-85.
- [8] L. Agapito, E. Hayman, I. Reid. Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision* 45 (2001) 107-127.
- [9] A. Senior, A. Hampapur, M. Lu. Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration. In *Proc. of Seventh IEEE Workshop on Applications of Computer Vision* (2002) 433-438.
- [10] J. Davis, X. Chen. Calibrating pan-tilt cameras in wide-area surveillance networks. In *Proc. of 9th IEEE International Conference on Computer Vision* (2003) 144-149.
- [11] H. Li, C. Shen. An LMI approach for reliable PTZ camera self-calibration. In: *Proc. of IEEE International Conference on Video and Signal Based Surveillance* (2006).
- [12] A. Basu, K. Ravi. Active camera calibration using pan, tilt and roll. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics* 27 (3) (1997) 559-566.

- [13] A. Ekin, A. M. Tekalp, R. Mehrotra: Automatic soccer video analysis and summarization. In: IEEE Transactions on Image Processing 12 (7) (2003).
- [14] K. Choi, Y.Seo. Tracking soccer ball in TV broadcast video. In CIARP 2005 LNCS 3617 (2005) 661–668.
- [15] Y. Ariki, S. Kubota, M. Kumano. Automatic production system of soccer sports video by digital camera work based on situation recognition. In Proceedings of ISM (2006) 851-860.
- [16] K. Kim, M. Grundmann, A. Shamir, I. Matthews, J. Hodgins, , I. Essa. Motion fields to predict play evolution in dynamic sport scenes. In Proceedings of CVPR (2010) 840–847.
- [17] W. Gomes, D. Borges. An algorithm for highlights identification and summarization of broadcast soccer videos. In CIARP 2012 LNCS (2012) 862–869.
- [18] E. Hayman, D. Murray. The effects of translational misalignment when self-calibration rotating and zooming cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (8) (2003) 1015–1020.
- [19] B. Triggs. Autocalibration from planar scenes. In Proceedings of the 5th European Conference on Computer Vision (ECCV’98) 1 , (1998) 89–105.
- [20] F. Szenberg, P. Carvalho, M. Gattass. Automatic camera calibration for image sequences of a football match. In Proceedings of ICAPR (2001) 2–8.
- [21] R. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (1986) 364–374.
- [22] K. Cornelis, N. Cornelis, M. Aerts , E. Ozden, L. Van Gool. Online calibration of two zoom-pan-tilt units for planar dynamic events. In Proceedings of the 28th conference on Pattern Recognition (DAGM’06) (2006) 698–707.

- [23] Y. Liu, Q. Huang, Q. Ye, W. Gao: 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).
- [24] M. Aleman-Flores, L. Alvarez, P. Henriquez, L. Mazorra: Morphological thick line center detection. In: 7th International Conference on Image Analysis and Recognition 6111 (2010) 71–80.
- [25] Q. Li and Y. Luo: Automatic camera calibration for images of soccer match. Proceedings of World Academy of Science, Engineering and Technology 1 (2005) 170–173.
- [26] D. Farin, S. Krabbe, PHN. de With, W. Effelsberg: Robust camera calibration for sport videos using court models. Storage and Retrieval Methods and Applications for Multimedia 5307 (2004) 80–91.
- [27] B. Jiang, L. Songyang and B. Liang: Automatic line mark recognition and its application in camera calibration in soccer video. IEEE International Conference on Multimedia and Expo (ICME) (2011) 1–6.
- [28] V. Babae-Kashany, H. Reza Pourreza: Camera pose estimation in soccer scenes based on vanishing points. 9th IEEE International Symposium on Haptic Audio-Visual Environments and Games (2010) 157–162.
- [29] L. Alvarez and J. Esclarin : Image quantization using reaction-diffusion equations. Siam Journal on Applied Mathematics 57 (1) (1997) 153–175.
- [30] T.H. Eagles: Constructive geometry of plane curves. MacMillan and co. 1885.
- [31] D.C. Brown. Close-range camera calibration. Photogrammetric Engineering 37 (1971) 855–866.
- [32] R. I. Hartley, A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press 2004.
- [33] C. McGlone. Manual of Photogrammetry, 4th ed. Amer. Soc. of Photogrammetry 1980.
- [34] F. Devernay, O. Faugeras. Straight lines have to be straight. Machine Vision and Applications 13 (1) (2001) 14–24.

- [35] S. Kang. Radial distortion snakes. In: Proc. IEICE Transactions on Information and Systems (2000) 1603–1611.
- [36] L. Alvarez, L.Gomez, J.R.Sendra. Algebraic lens distortion model estimation. In: Image Processing On line (2010) http://www.ipol.im/pub/demo/ags_algebraic_lens_distortion_estimation.
- [37] L. Alvarez and V. Caselles. Calibration method for a TV and video camera. European Patent 09380137.1. Issued February 2011.