Robust Camera Calibration of Soccer Video using Genetic Algorithm

Reza Pourreza† Morteza Khademi† Hamidreza Pourreza‡ Habib Rajabi Mashhadi†

reza_pourreza@yahoo.com khademi@um.ac.ir hpourreza@um.ac.ir h_mashhadi@um.ac.ir

†. Electrical Engineering Department, Ferdowsi University of Mashhad

‡. Computer Engineering Department, Ferdowsi University of Mashhad

Abstract

Abstract- This paper proposes an exact and robust genetic algorithm-based method for the calibration of soccer camera. According to the FIFA official soccer field layout we defined a field model for the soccer court. Camera calibration is done through finding the homography transform between the field model and the input frame which is followed with DLT decomposition. The intersection of lines in the field model and input frame form feature points and by means of a genetic algorithm, we found the correspondence between those points. Our algorithm was applied to a couple of soccer video frames and the achieved results demonstrate its robustness, accuracy and high performance.

1. Introduction

Automatic sports video analysis is an emerging topic in the field of computer vision, since manually analyzing of the lengthy sports video is a time-consuming and exhaustive process. Many researchers have invested in this field and their researches' topics involve *sport event detection*, *automatic sport video retrieval*, *augmented reality*, *virtual advertisement* and *referee assistant*. Soccer is a spectacular sport and has a lot of viewers around the world, so more investment is needed in comparison with other sports.

Camera calibration is a building block of many automatic sports video analysis systems and its applications include *ball and players' position detection*, *3D tracking of ball and players*, *3D reconstruction of video*, *augmented reality* and *virtual advertisements*. In order to calibrate the camera some reference points should be extracted from the given video frame. If the coordinates of reference points in the world coordinates system are known, camera calibration is done by solving an equation, whereas the correspondence between reference points in image coordinates and world coordinates systems must be found if the coordinates are unknown. In [1]-[4], the correspondence is assumed known, thus one should find the correspondence manually and so the systems are not fully automatic. In [5] and [6], the systems search over the whole camera parameter space to find the best parameters, these systems are automatic but they do an exhaustive search and are computationally expensive. In [7]-[9], the correspondence is assumed unknown and the systems check all possible correspondences to find the best one. The last systems are more efficient than the previous ones; however, they perform a full search and are expensive.

The aim of this paper is to find the soccer camera parameters when a frame of broadcasted soccer video is given. We have proposed a method which uses a genetic algorithm to find the optimum camera parameters instead of doing an exhaustive full search over the parameter space. The rest of paper is organized in four sections: *system overview*, *preprocessing*, *genetic algorithm implementation*, *results and discussion*.

2. System Overview

Pinhole camera model consists of 11 parameters including camera position (X_c, Y_c, Z_c), camera orientation (α, β, γ), principal point coordinates (U_p, V_p), focal length (f) and skewness coefficients (b_1, b_2). The Direct Linear Transform decomposition (DLT) method is used to find the camera parameters. The input of DLT decomposition algorithm is the homography transform between the field model in the world coordinates system and the captured field in the image coordinates system. A model is defined for soccer field as a source for extracting feature points in the world coordinates system. The intersection of field

978-1-4244-2673-7/08/$25.00 ©2008 IEEE

lines in the world and image coordinates systems are chosen as feature points to obtain the homography transform. Feature points in the image coordinates are extracted from input frame and feature points in world coordinates are extracted from the field model. FIFA standard layout for soccer field and the chosen model for calibration are depicted in fig. 1.

Since the feature points in the world coordinates are coplanar, the homography transform matrix (H) dimension is 3x3 with an element (h_{22}) for normality and its degree of freedom is limited to 8. If four non-colinear feature points in the world coordinates system $(X_i:(x_i, y_i, 1), i = 1,2,3,4)$ and four corresponding feature points in the image coordinates system $(U_i:(u_i, v_i), i = 1,2,3,4)$ exist, H can be computed using (1).

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1v_1 & -y_1v_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2u_2 & -y_2u_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2v_2 & -y_2v_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3u_3 & -y_3u_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3v_3 & -y_3v_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4u_4 & -y_4u_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4v_4 & -y_4v_4 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \end{bmatrix} \quad (1)$$

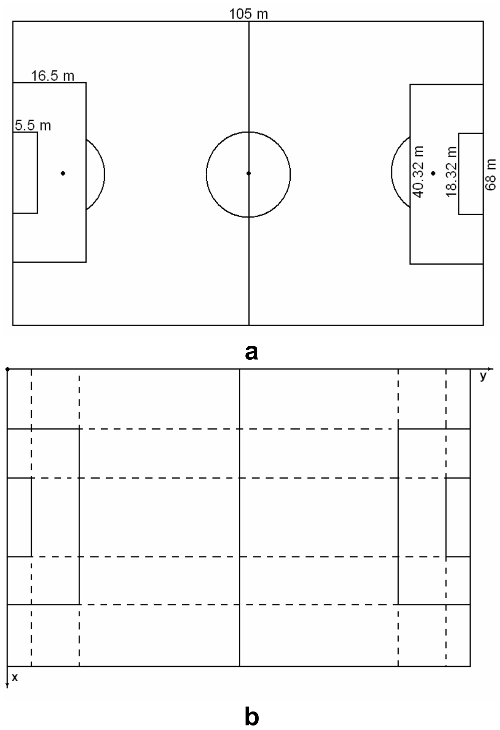

Fig. 1. Soccer Field Models
a. FIFA Standard Model
b. Defined Model for Calibration

The H decomposition process using DLT algorithm to achieve camera parameters is available in details in [10]. However, the following two assumptions must be considered in the process of H decomposition, since the degree of freedom of H is 8 and the number of unknown camera parameters is 11:

1. The coordinates of principal point is known to be the frame center, $(H/2, W/2)$.
2. b_1 is known and equals 1.

Therefore, the problem of calibration is reduced to find the homography matrix. As annotated earlier, H can be computed, if four feature points in the image coordinates and four corresponding feature points in the world coordinates are available. Feature points in the world coordinates are the intersections of field lines which their exact positions are known. In order to extract them, the field lines are divided into two groups, horizontal and vertical lines. If two vertical and two horizontal lines are selected and their intersections points are computed, four feature points in the world coordinates would be acquired $(X_i, i = 1,2,3,4)$.

In order to extract the feature points in the image coordinates, firstly field lines directions in the input frame should be extracted which is completely explained in *preprocessing* section. The extracted field lines are divided into two groups, horizontal and vertical lines. The criteria for assigning a line to a group is its direction in Hough space (θ), the lines with -20°<θ<20° are labeled horizontal and other lines are vertical. If two vertical and two horizontal lines are selected and their intersections points are computed, four feature points in the world coordinates would be acquired $(U_i, i = 1,2,3,4)$. These feature points can lie either inside or outside the frame.

H matrix can be computed using (1), X_i and U_i. However one can not claim that the computed H is true and optimum, because we do not now whether the two 4-feature-points are in correspondence. In order to find the correspondence between feature points to achieve the optimum H, a genetic algorithm is used which finds the best matching between lines intersection points. The details of our algorithm are available in *Genetic Algorithm Implementation* section.

3. Preprocessing

The purpose of preprocessing is to extract the field lines in the input frame. Extraction of field lines is a two-step process, in the first step the grass area is detected and in the second step white lines are extracted from the detected grass area. This strategy ensures the system that the extracted lines are field lines and other present lines in the frame are ignored.

Detection of grass area is based on the histogram of Hue component in the HSI color space, because the Hue component contains color information. For this purpose a 256-bin histogram is computed. The histogram peak index (i_p) and an interval around it $([i_{\min}, i_{\max}])$ are considered as the grass hue bins. $i_{\min}$ and $i_{\max}$ are computed using (2) to (7). In these equations K is a constant and equals 0.15 and $Hist$ denotes Hue histogram.

$$Hist[i_{\min}] \geq k \times Hist[i_p] \quad (2)$$

$$Hist[i_{\min} - 1] < k \times Hist[i_p] \quad (3)$$

$$Hist[i_{\max}] \geq k \times Hist[i_p] \quad (4)$$

$$Hist[i_{\max} + 1] < k \times Hist[i_p] \quad (5)$$

$$i_{\min} \leq i_{peak} \quad (6)$$

$$i_{\max} \geq i_{peak} \quad (7)$$

The input frame and detected grass area shown in fig. 2 and 3 respectively. The black pixels in fig. 3 denote grass pixels.

In order to extract field lines in the grass area, firstly the detected grass area is changed to obtain a uniform region. For this purpose, a median filter and a morphological Opening operation are applied to the detected grass area. The obtained uniform region for grass area is shown in fig. 4.

Fig. 4- Uniform Grass Area

Blue component of the input frame in RGB space is chosen for white-pixel detection and edge detection operations which are applied to the uniform grass area. The reason of choosing B component is the high contrast between green grass pixels and white line pixels in this component.

Sobel operator is used for edge detection. The criterion for a pixel to be considered as white is its gray level of B component that should exceed 150.

The combination of white-pixel and edge detection gives good candidate pixels for field lines. After applying a morphological Thinning operation to the combination, a standard Hough transform is used to find the directions of field lines. The output of thinning operation and the detected field lines directions are shown in fig. 5 and fig. 6 respectively.

By the end of field line detection, vertical and horizontal lines are separated based on their angle in

Fig. 2- Input frame

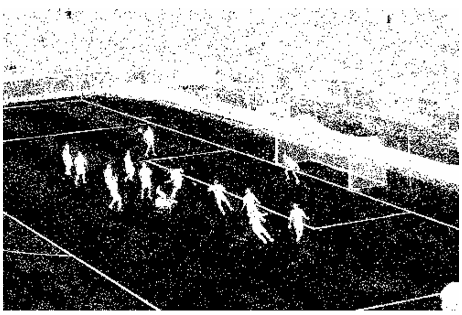

Fig. 3- Detected Grass Area

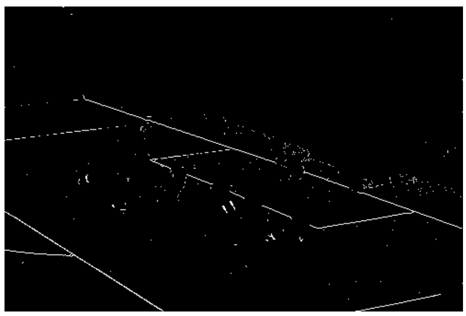

Fig. 5- Detected Line Pixels

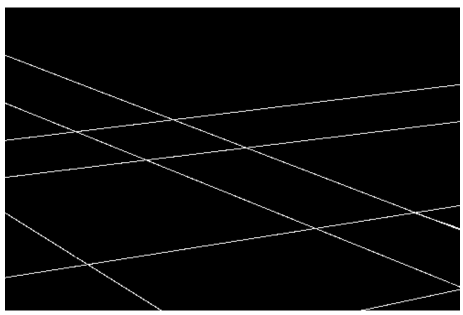

Fig. 6- Detected Line Directions

Hough space and then are sorted and labeled. Therefore two groups of lines are obtained, one is the group of horizontal lines which has *nr_hor* entries and the other is the group of vertical lines which has *nr_ver* entries. These two groups are the input of genetic algorithm to find the best matching between intersection points.

4. Genetic Algorithm Implementation

The aim of utilizing genetic algorithm is to find the optimum H matrix through finding the best correspondence between feature points. For this purpose one should pursue the following procedure:

1. Select two vertical and two horizontal lines among the lines in the field model and find the intersection points.
2. Select two vertical and two horizontal lines among the lines in the input frame and find the intersection points.
3. Compute H matrix using equation (1).
4. Transform the field model to the image coordinates by the obtained H.
5. Score the matching between transformed model (*transLines*) and the extracted field lines directions in the input image (*frameLineDirs*).
6. Repeat steps 1 to 5 for all possible choices and find the maximum score.

Our approach is like above procedure, however, instead of performing an exhausting full search among all possible feature points, a genetic algorithm is utilized to perform an intelligent search to find the best correspondence. The genetic algorithm selects appropriate vertical and horizontal lines automatically, evaluates the obtained H and scores the correspondence. Characteristics of the utilized GA are explained in subsections *a* to *h*.

a. Chromosome: Bit string

A	B	C	D	E	F	G	H
3 bits	3 bits	3 bits	3 bits	3 bits	3 bits	3 bits	3 bits

A:	Model's 1st Horizontal Line Index	*1~6*
B:	Model's 2nd Horizontal Line Index	*1~5*
C:	Model's 1st Vertical Line Index	*1~7*
D:	Model's 2nd Vertical Line Index	*1~6*
E:	Image's 1st Horizontal Line Index	*1~nr_hor*
F:	Image's 2nd Horizontal Line Index	*1~nr_hor−1*
G:	Image's 1st Vertical Line Index	*1~nr_ver*
H:	Image's 2nd Vertical Line Index	*1~nr_ver−1*

b. Fitness: In order to evaluate fitness this important rule "Each line in the field model coordinates corresponds with only one line in input frame" is considered. Fitness evaluation procedure is as follow:

For i = 1 to 13
    Transform i^{th} field line to obtain *transLine*
    Evaluate i^{th} line fitness, $fitness_i$ using (8)
    Update *frameLineDirs* using (9)
Loop

$$\text{fitness} = \sum_{i=1}^{13} \text{fitness}_i$$

$$fitness_i = \begin{cases} \sum_{i=1}^{H}\sum_{j=1}^{W} f(i,j) & ,\sum_{i=1}^{H}\sum_{j=1}^{W} f(i,j) > 100 \\ -0.5\times\sum_{i=1}^{H}\sum_{j=1}^{W} f(i,j) & ,otherwise \end{cases} \quad (8)$$

$$f(i,j) = \begin{cases} 1 & ,if\ transLine(i,j) = 1 and \\ & \quad frameLineDirs(i,j) = 1 \\ 0 & ,otherwise \end{cases}$$

$$frameLineDirs = frameLineDirs \,\&\, (not\ transLine) \quad (9)$$

c. Selection: Roulette Wheel

d. Crossover: Single point with the probability of 0.8

e. Mutation: Mutation rate is a function of current generation index.
mutate_rate=(10-mod(generation_index,10))×0.05

f. Elitism: Keep the best individual

g. Generation Size: 50

h. Maximum No. of Generation: 500

5. Results and Discussion

We have applied the proposed GA-based algorithm to several soccer video frames which included the goal mouth. Our algorithm is as exact as the full search algorithms but gets to the optimum point faster. An instance of fitness evaluation through 500 generations for a sample frame is shown in fig. 7. The maximum fitness through all generations equals 2487.

Two vertical and horizontal lines in input frame and field model which made the maximum fitness were selected and their corresponding H matrix was computed. The field model was transformed to the image coordinates system using the optimum H. Fig. 8 shows the input frame and the transformed field model is overlaid. This figure demonstrates best matching between transformed field model and frame's field lines.

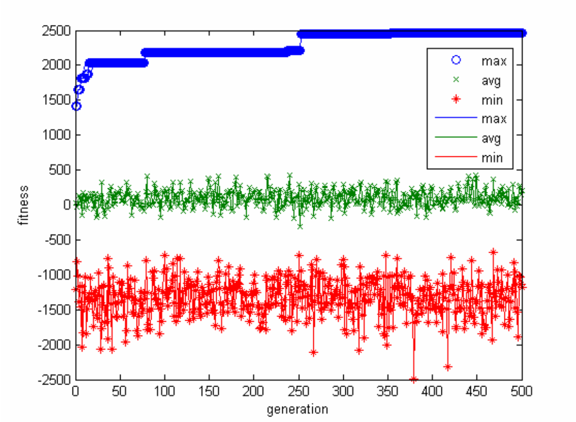

Fig. 7- Best, Mean and Worst Fitness through all Generations

Fig. 8- Best matching

6. References

[1] Yang Liu, Qingming Huang, Qixiang Ye, Wen Gao, "A New Method to Calculate the Camera Focusing Area and Player Position on Playfield in Soccer Video", in *Proc of SPIE*, Vol. 5960, 2005.

[2] Yang Liu, Dawei Liang, Qingming Huang, Wen Gao, "Extracting 3D information from broadcast soccer video", *Image and Vision Computing*, Vol. 24, 2006, pp. 1146–1162.

[3] Dawei Liang, Qingming Huang, Yang Liu, Guangyu Zhu, Wen Gao, "Video2Cartoon: A System for Converting Broadcast Soccer Video into 3D Cartoon Animation", *IEEE Transactions on Consumer Electronics*, Vol. 53, No. 3, 2007.

[4] Yang Liu, Dawei Liang, Qingming Huang, Wen Gao, "Self-calibration Based 3D Information Extraction and Application in Broadcast Soccer Video", *ACCV2006*, pp. 852–861.

[5] T. Watanabe, M. Haseyama, H. Kitajima, "A soccer field tracking method with wire frame model from TV images", in *Proc of International Conference on Image Processing*, Vol. 3, 2004, pp. 1633-1636.

[6] Akihito Yamada, Yoshiaki Shirai, Jun Miura, "Tracking Players and a Ball in Video Image Sequence and Estimating Camera Parameters for 3D Interpretation of Soccer Games", in *Proc of International Conference on Pattern Recognition*, Vol. 1, 2002.

[7] Flávio Szenberg, Paulo Cezar Pinto Carvalho, Marcelo Gattass, "Automatic Camera Calibration for Image Sequences of a Football Match", in *Proc of the Second International Conference on Advances in Pattern Recognition*, 2001.

[8] Dirk Farin, Jungong Han, and Peter H. N. de With, "FAST CAMERA CALIBRATION FOR THE ANALYSIS OF SPORT SEQUENCES", in *Proc of IEEE International Conference on Multimedia and Expo*, 2005.

[9] Dirk Farin, Susanne Krabbe, Peter H.N. de With, Wolfgang Effelsberg, "Robust Camera Calibration for Sport Videos Using Court Models", in *Proc. of SPIE Storage and Retrieval Methods and Applications for Multimedia*, vol. 5307, 2004, pp. 80-91.

[10] T. Melen, "Extracting Physical Camera Parameters from the 3x3 Direct Linear Transformation Matrix", in *Proc. of SPIE*, Vol. 2252, 1994, pp. 355-365.