Camera Pose Estimation in Soccer Scenes Based on Vanishing Points

Vahid Babaee-Kashany
Machine Vision Research Laboratory
Ferdowsi University of Mashhad
Mashhad, Iran
v.babaee@gmail.com

Hamid Reza Pourreza
Machine Vision Research Laboratory
Ferdowsi University of Mashhad
Mashhad, Iran
hpourreza@um.ac.ir

Abstract— **Camera parameters estimation is an important issue in machine vision. This paper proposes a new method to find translation and rotation matrix of camera in sport scene on the basis of vanishing points. Vanishing point (VP) of parallel lines is the image of the point at infinity, which corresponds to the projection of the intersection of parallel lines at infinity. According to projective geometry constraint, camera rotation of the projection matrix is computed directly by two vanishing points and pan and tilt of camera extracted from this matrix. Mathematical proof and Computer simulation are carried out to validate our novel method.**

Keywords- Camera calibration; pose estimation; Rotation matrix; vanishing points.

I. INTRODUCTION

Soccer is the most popular sport in the world with tremendous amount of video programs produced every year. Automatically analysing soccer videos, such as finding some exciting events for summarizing, is a hot research area which will help professionals to analyse teams' tactics, strengths and weaknesses. Many other researchers have investigated in this field and their researches' topics involve sport event detection, automatic sport video retrieval, augmented reality, virtual advertisement and referee assistant.

In order to calibrate camera, some reference points should be extracted from the given video frame. If the coordinate systems are known, camera calibration is done by solving an equation, whereas the correspondence between reference points in the image coordinates system must be found if coordinates are known.

The aim of this paper is to find dynamic camera parameters when a frame of soccer video is given. We proposed a method which uses geometry of soccer model to find the rotation matrix instead of doing an exhaustive full search over parameter space to find all camera parameters and decompose it.

In recent years, some researchers used correspond between lines to camera calibration. Yu and Jiang [1] proposed an offline method to find external and internal parameters on the basis of frame grouping and Hough like search. Farin et.al. [2], Battikh and Jabro [3] used court model and Kanade-Lucas-Tomasi (KLT) tracker to release fully automatic method, the last one used a hardware accelerator to achieve real time video content insertion.

Kim and Hong [4] also propose a calibration algorithm for soccer games based on a pan-tilt camera (without roll) but still the inter-frames transformation are estimated by identifying corresponding line between frames and using a non-linear approach to determine the homography matrix that minimizes the Euclidean distance between line pairs.

All these different methods for calculating the camera parameters and focusing area on soccer scenes have a similar characteristic, i.e., they require a number of corresponding points (at least 4 points) to determine the homography matrix for an image. Since the main camera of soccer playfield is free in panning, tilting and zooming, it cannot be guaranteed that every image has sufficient corresponding points. But here we assume that there is a fix camera position and also find variable camera parameters (pan and tilt) without any correspondence.

The rest of paper is organized in four sections: Geometry of vanishing point, Camera calibration using vanishing points, Static Camera Parameters, Experiments and Results.

II. GEOMETRY OF VANISHING POINT

A. *Camera Model and the Concept of Vanishing Point*

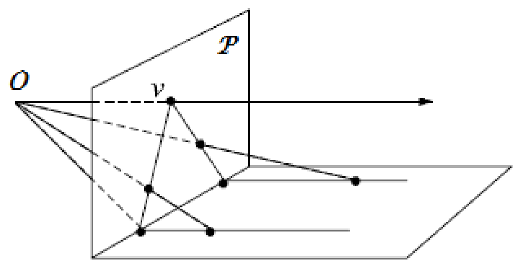


Fig 1. The line connecting the vanishing point with the optical center has the same direction with the corresponding lines in 3D space.

For the classic pinhole model, the basic formula of perspective projection is given by:

$$\lambda_m .m = K.[R \quad T].M \tag{1}$$

Where:

M denotes a 3D point and m denotes the corresponding 2D point on image. They are both expressed in homogeneous coordinate and λ_m is an arbitrary scale factor.

R is 3 x 3 rotation matrix that describes the rotational mapping from the world coordinate system into the camera coordinate system.

T is a 3 x 1 vector that describes the translational mapping from the world coordinate system into the camera coordinate system.

K is a 3 x 3 matrix describing the internal camera parameters,

$$K = \begin{bmatrix} f & s & u_0 \\ 0 & \beta.f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Where f is scale factor in image u- and v- axes, s is the parameter describing the skew of the two images axes, Beta is non-isotropic scaling and (u_0, v_0) are the coordinates of the principal point.

As in Fig 1, for the perspective projection, the images of parallel lines meet at a point, if they are not parallel to the image plane. This point is called vanishing point[5]. The line connecting the vanishing point with the optical center has the same direction with the corresponding line in 3D space [6]. That means all parallel lines in 3D space correspond to the same vanishing point in image space.

B. Geometry of Vanishing Point in Soccer Scene

As shown in Fig 2(a), the court model is made of two groups of parallel lines: vertical lines and horizontal lines. We assume that the origin of the world coordinate system coincides with point D and suppose l_1 and l_2 are Length and width of biggest rectangle respectively, and then the homogeneous coordinates of vertexes A, B, C and D are described as:

$$\begin{array}{ll} X_a = (0, l_2, 0, 1)^T & X_b = (l_1, l_2, 0, 1)^T \\ X_c = (l_1, 0, 0, 1)^T & X_d = (0, 0, 0, 1)^T \end{array} \quad (3)$$

The projective image of the court model is shown as in Fig 2(b). v_1 and v_2 correspond to vanishing points of axes X and Y respectively. The perspective relationship between v_1, v_2 and the corresponding 3D points are [5]:

$$\lambda_1.v_1 = K[R \quad T]\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = K[r_1 \quad r_2 \quad r_3 \quad T]\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = K.r_1 \quad (4)$$

$$\lambda_2.v_2 = K[R \quad T]\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = K[r_1 \quad r_2 \quad r_3 \quad T]\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = K.r_2 \quad (5)$$

According to equations (4) and (5), the vanishing points are Independent from Length and width of rectangle.

Because the rotation matrix R is unitary and orthogonal [7], then we have

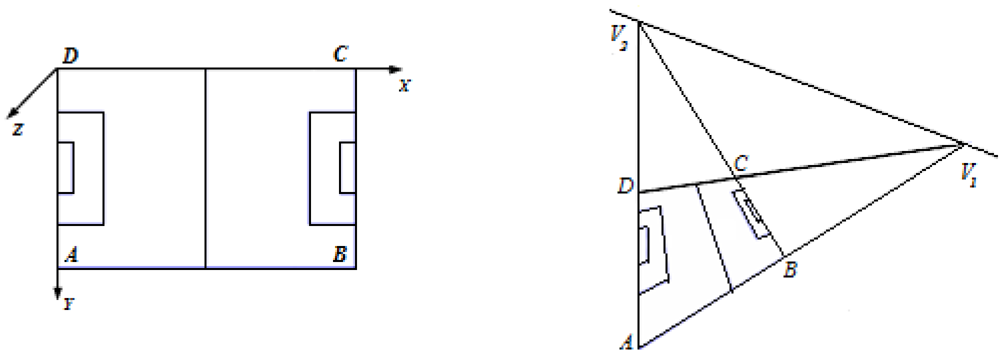
$$\begin{cases} r_i^T.r_j = 0 \\ \|r_i\| = \|r_j\| = 1 \end{cases} \quad (i \neq j) \quad (6)$$

From (4) to (6), the following equations are satisfied:

$$\lambda_2 K^{-T} v_2^T \cdot \lambda_1 K^{-1} v_1 = 0 \quad (7)$$

$$\|\lambda_1.K^{-1}.v_1\| = \|\lambda_2.K^{-1}.v_2\| = 1 \quad (8)$$

According to (6) and (7), the vectors $\lambda_1.K^{-1}.v_1$, $\lambda_2.K^{-1}.v_2$ are mutually orthogonal and have an equal modulus.



(a) (b)

Fig 2. (a) Court model, (b) is perspective image in world coordinate system.

As already noted, the line that connects the vanishing point with the optical center, has the same direction with the corresponding line in 3D space and as shown in Fig 3, two vectors Ov_1 and Ov_2 are perpendicular. Thus, these two vectors that is created by two vanishing points, can build a parallel plane to soccer play-field.

III. CAMERA CALIBRATION USING VANISHING POINTS

A. Using Vanishing Point

Assume $P_{1\infty}$, $P_{2\infty}$ and $P_{3\infty}$ are three infinity point in direction of x, y and z axis of camera that created with a pair of orthogonal and parallel line(s) in the 3D space, and the corresponding vanishing points are p_1, p_2 and p_3 on the image plane. According to pinhole camera model, the relationship between the infinity points and their image projection is given by:

$$\begin{cases} \lambda_1 p_1 = K[R \quad T]P_{1\infty} \\ \lambda_2 p_2 = K[R \quad T]P_{2\infty} \\ \lambda_3 p_3 = K[R \quad T]P_{3\infty} \end{cases} \tag{9}$$

So from (1), (4) and (5) and under the assumption of known zero skew, the equation (9) can be rewritten as follows:

$$\begin{bmatrix} \lambda_1 u_1 & \lambda_2 u_2 & \lambda_3 u_3 \\ \lambda_1 v_2 & \lambda_2 v_2 & \lambda_3 v_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & \beta.f & v_0 \\ 0 & 0 & 1 \end{bmatrix} R \tag{10}$$

Where, (u_1, v_1) is vanishing point of vertical line and (u_2, v_2) is vanishing point of horizontal lines, thus

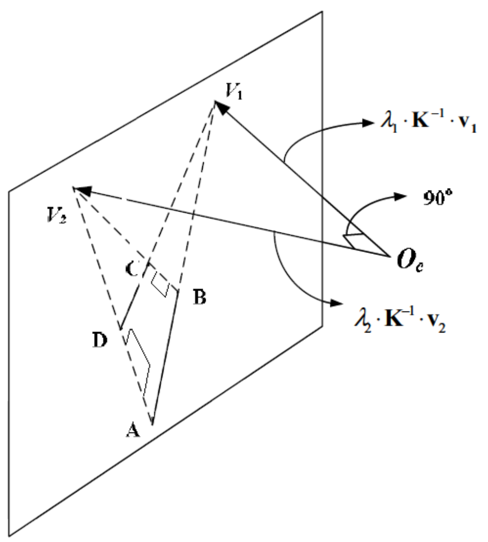


Fig 3. Geometric relationship between vanishing points in the camera coordinates system in sport scene.

$$R = \frac{1}{f\beta} \begin{bmatrix} \lambda_1\beta(u_1 - u_0) & \lambda_2\beta(u_2 - u_0) & \lambda_3\beta(u_3 - u_0) \\ \lambda_1(v_1 - v_0) & \lambda_2(v_2 - v_0) & \lambda_3(v_3 - v_0) \\ f\beta\lambda_1 & f\beta\lambda_2 & f\beta\lambda_3 \end{bmatrix} \tag{11}$$

There are five unknown parameters and two vanishing points, so we must find some more equations in order to obtain the solution (we have focal length, non-isotropic scaling and principal point).

B. Obtaining Third Vanishing Point

According to relationship between vanishing vectors in 3D space, we can find third vanishing point from cross product of two other vanishing vectors [5].

The orthonormality of R(6) can be used to provide the following equation from (11):

$$\beta(u_1 - u_0)(u_2 - u_0) + (v_1 - v_0)(v_2 - v_0) + \beta = 0 \tag{12}$$

$$\beta(u_1 - u_0)(u_3 - u_0) + (v_1 - v_0)(v_3 - v_0) + \beta = 0 \tag{13}$$

$$\beta(u_2 - u_0)(u_3 - u_0) + (v_2 - v_0)(v_3 - v_0) + \beta = 0 \tag{14}$$

Subtracting (14) from (12) and (13) gives:

$$\begin{cases} \beta(u_1 - u_0)(u_2 - u_3) + (v_1 - v_0)(v_2 - v_3) = 0 \\ \beta(u_2 - u_0)(u_1 - u_3) + (v_2 - v_0)(v_1 - v_3) = 0 \end{cases} \tag{15}$$

So, the third vanishing point is recovered.

C. Obtaining λ_i

In order to obtain a geometry interpretation of λ_i , row normality must be considered[8]. This gives:

$$\begin{cases} \left(\lambda_1^2(u_1 - u_0) + \lambda_2^2(u_2 - u_0) + \lambda_3^2(u_3 - u_0)\right)/ f = 0 \\ \left(\lambda_1^2(v_1 - v_0) + \lambda_2^2(v_2 - v_0) + \lambda_3^2(v_3 - v_0)\right)/ \beta.f = 0 \\ \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 1 \end{cases} \tag{16}$$

Subtracting third equation from two other one, gives:

$$\begin{cases} \lambda_1^2 = \dfrac{(v_0 - v_3)(u_2 - u_3) - (u_0 - u_3)(v_2 - v_3)}{(v_1 - v_3)(u_2 - u_3) - (u_1 - u_3)(v_2 - v_3)} \\ \lambda_2^2 = \dfrac{(v_1 - v_3)(u_0 - u_3) - (u_1 - u_3)(v_0 - v_3)}{(v_1 - v_3)(u_2 - u_3) - (u_1 - u_3)(v_2 - v_3)} \\ \lambda_3^2 = 1 - \lambda_1^2 + \lambda_2^2 \end{cases} \tag{17}$$

So, the λ_i is recovered.

D. Recovery of the Rotation Matrix

Now, in order to calculate camera rotation on world coordinate system, we placed lambdas and third vanishing point from (17) and (15) to equation (11). By decomposition of rotation matrix we can obtain rotation of camera, on camera coordinate system. Of course, this matrix actually describes a rotation of the world coordinate system (rather than the camera). But camera rotation matrix is equivalent to an inverse rotation of the world [9].

IV. STATIC CAMERA PARAMETERS

In soccer video some of prior knowledge can be used to simplify the process of camera calibration. First, the playfield is a plane, which reduce 3*4 camera matrix to 3*3 one. Second, the position of main camera in soccer playfield is almost fixed and we have considered situation which the focal length is also fixed. Here the homography between playfield model and its image is used to determine the 3D position of main camera in world coordinate system. Third, according to FIFA's laws of the game, the lengths of lines within penalty area are known and it can be used to estimate homography matrix (Fig 4).

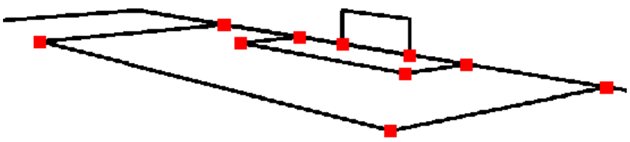


Fig 4. The point marked by dots can be used to estimate homography matrix.

In particular, let $[x, y]$ denote the 2D position of a point on a video frame and its 3D position on the playfield could be denoted as [X, Y, Z]. Under above assumptions, the Z values are a linear function of X and Y. Thus, the relationship between the image coordinates and the 3D coordinates can be formulated as 3 x 3 homography matrix [10]:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{18}$$

Where $h_{33}=1$, which is always true for model to image transformation.

A. Recovery of the Camera Position

Now the camera position is readily computed. The camera's 3D position t_{cw} in world coordinate system can be obtained by [11]:

$$t_{cw} = -\begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix}^{-1} \begin{bmatrix} t_x & t_y & t_z \end{bmatrix} \tag{19}$$

Where

$$\begin{bmatrix} t_x & t_y & t_z \end{bmatrix} = \frac{2K^{-1}h_3}{\left\| K^{-1}h_1 \right\| + \left\| K^{-1}h_2 \right\|} \tag{20}$$

$$r_1 = \frac{K^{-1}h_1}{\left\| K^{-1}h_1 \right\|},\ r_2 = \frac{K^{-1}h_2}{\left\| K^{-1}h_2 \right\|} \tag{21}$$

and r_3 can be found by cross product of r_1 and r_2 [10].

B. Recovery of the Focal length

The Equation 18 has eight degree of freedom but the real-world image formation process has only seven. These comprise one for focal length, three for camera rotation and three for camera position. The eighth degree can be attributed to non-isotropic scaling which refer to unequal scaling in x and y axis direction in video frame. To determine non-isotropic scaling from H, we first compensate the camera principal point (u_0, v_0) which we assume to be at the center of the image by multiplying a homography matrix to the left side[2].

$$\mathrm{H}' = \begin{bmatrix} h'_{11} & h'_{12} & h'_{13} \\ h'_{21} & h'_{22} & h'_{23} \\ h'_{31} & h'_{32} & h'_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{H} \tag{22}$$

hence

$$\mathrm{H}' = \begin{bmatrix} fr_{11} & \beta fr_{12} & ft_x \\ fr_{21} & \beta fr_{22} & ft_y \\ r_{31} & \beta r_{32} & t_z \end{bmatrix} \tag{23}$$

Since the rotation matrix is Orthonormal, f and β are calculated with the following formulas:

$$f = \sqrt{-\frac{h'_{11}h'_{12} + h'_{21}h'_{22}}{h'_{31}h'_{32}}} \tag{24}$$

$$\beta = \sqrt{\frac{h'^2_{12} + h'^2_{22} + f^2 h'^2_{32}}{h'^2_{11} + h'^2_{21} + f^2 h'^2_{31}}} \tag{25}$$

V. EXPERIMENTS AND RESULTS

A. Simulations

The reliability and validity of our proposed approach has been tested by computer simulation data. The simulator camera parameter characterized by following properties: s=0, β =0, $(u_0, v_0) = (400{,}300)$ and the image resolution is 800×600. Our model parameters are l_1=10700cm and l_2=7400cm. In simulations, we fixed camera position in

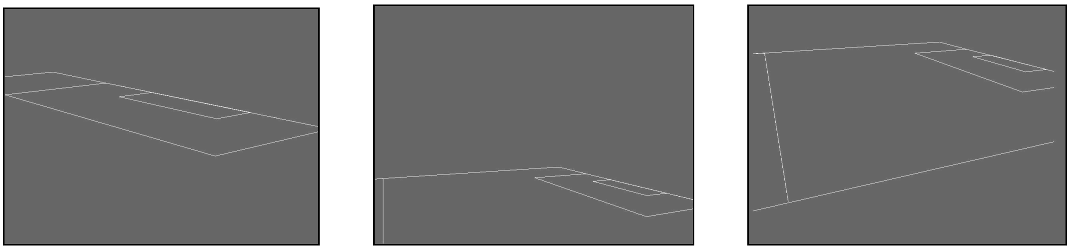


Fig 5. Three image that captured from simulator

$r_1 = [39.6° \quad -12.6°],\ f_1 = 1600;\ r_2 = [37.8° \quad -3.6°],\ f_2 = 800;\ r_3 = [23.4° \quad -16°],\ f_3 = 800$

$T_c = [5350, 9620, -1440]^T$ and then changed camera pan and camera tilt (roll=zero).

As shown in Fig 5, we used three Images that are captured from computer simulator with various focal length and camera rotation angles. Then we found vanishing points and calculated rotation matrix and estimated pan and tilt of camera.

B. Finding vanishing points

The first step in camera calibration to recovery projection matrix is finding vanishing points.

For this purpose the Hough transform is used. Considering that the angle of horizontal lines in the main stadium camera always is between zero and 35 degree and the angle of vertical lines between 55 to ninety degrees, Hough transform selects two lines from among vertical lines and two lines among horizontal lines. Then by these two categories of parallel lines, vanishing points were calculated.

C. Rotation matrices

Finding the vanishing points, is the second step to recover the camera rotation matrix from (11), when (u_3, v_3) and λ_i are obtained by (15) and (17). To estimate the camera pan, tilt and roll, we use inverse decomposition of rotation matrix,

D. Simulation Experiments

On the court model, 15% of points of each line are dropped. Gaussian noise with mean 0 and standard deviation ranging 0 to 4 pixels is added to each projected image point. For each noise level, we performed 100 runs and the results shown are

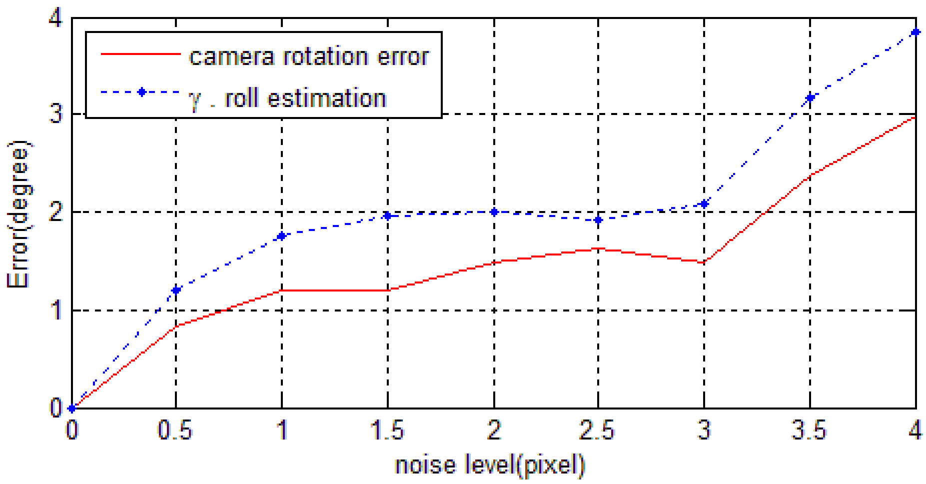


Fig 6. Relation between rotation error and roll estimation (we know roll must be zero) at different noise levels.

TAB 1. RESULT OF MAXIMUM NUMBER OF PROJECTION AND ACCURACY OF OUR METHOD AND CORRESPONDENCE METHOD.

Best Camera Parameters	**Our new method (frames with roll < 1 degree)**	**Correspondence method**
Number of projection (fix focal length)	36	1080
Number of projection (variable focal length)	36 (f_{max}-f_{min})	1080
Accuracy of Projection	2 pixel	Depended on quality of line detection

average.

Fig 6 gives relationship between camera rotation error and roll estimation. According to our experimental results, pan and tilt estimation error have direct relationship with roll estimation, as follow:

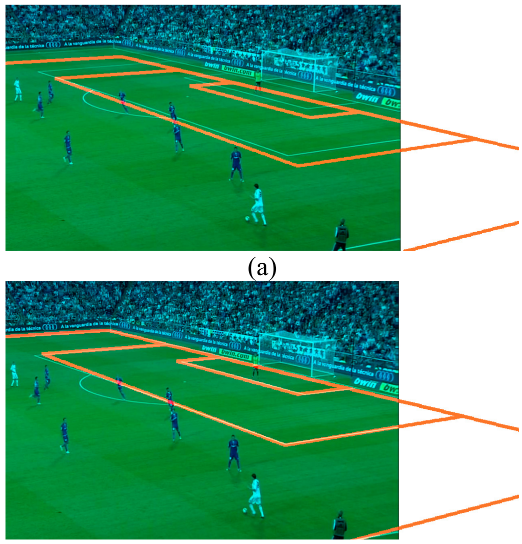
$$e_{pan} + e_{tilt} \leq \gamma \times roll \tag{26}$$

Where, γ is a threshold factor that set up to 4 (however it can be set at a lower number but we get best result with this value).

The results of our new method and correspondence method are shown in Tab 1 ($f_{\min}$ and $f_{\max}$ are minimum and maximum of camera focal length). Here the first and second rows describe the Max number of Model projection for find best camera parameters in fixed focal length and variable focal length. The Accuracy of projection is shown from third row in table.

E. Real data Experiments

We have tested the algorithm on 500 fixed focal length frames that recorded from regular television broadcasts. In first step, we calculated camera static parameters using ten frames containing adequate feature points, and then used (19), (24) and (25) to achieve camera position, focal length and non-isotropic scaling, respectively. In second step, we calculated camera pan, tilt and roll using vanishing points. As already noted, lines in real videos are not accurately obtainable and as shown in Fig 7 court model cannot project to frame with this noisy parameters. Thus, by using (26) error in pan and tilt estimation is reduced. For further research we can use this method to achieve High-accuracy virtual content insertion in long sequence of frames by a feature tracker.



(a)

(b)

Fig 7. (a) Is a sample frame before noise reduction (pan= 27.2°, tilt= -8.51°, roll= 0.81°), (b) is the corrected version of (a) by using equation (26) (pan= 28.02°, tilt= -9.11°, roll= 0°) ,

Camera static parameters : T_{cw}=[6146, 9085, -96.04], f=1801, B=0.9226 .

ACKNOWLEDGMENT

We would like to thanks Mr. Xiaolong Zhou and Dr. En Peng for their valuable comments about the vanishing point approach.

REFERENCES

[1] X. Yu, N. Jiang, L.-F. Cheong, H. W. Leong, and X. Yan, "Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking," *Computer Vision and Image Understanding,* vol. 113, pp. 643-652, 2009.

[2] D. Farin, S. Krabbe, P. H. N. de With, and W. Effelsberg, "Robust camera calibration for sport videos using court models," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, San Jose, CA, USA, 2003, pp. 80-91.

[3] T. Battikh and I. Jabri, "Camera calibration using court models for real-time augmenting soccer scenes," *Multimedia Tools and Applications,* pp. 1-15, 2010.

[4] H. Kim and K. Sang Hong, "Robust Image Mosaicing of Soccer Videos using Self-Calibration and Line Tracking," *Pattern Analysis & Applications,* vol. 4, pp. 9-19, 2001.

[5] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer Vision,* vol. 4, pp. 127-139, 1990.

[6] S. Jin, G. Hongbin, Q. Xiaolin, and Z. Jiping, "A new camera calibration based on vanishing point," in *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, 2008, pp. 2371-2376.

[7] H. Bacakoglu and M. S. Kamel, "A three-step camera calibration method," *Instrumentation and Measurement, IEEE Transactions on,* vol. 46, pp. 1165 -1172, 1997.

[8] B. W. He and Y. F. Li, "Camera calibration from vanishing points in a vision system," *Optics & Laser Technology,* vol. 40, pp. 555-561, 2008.

[9] A. C. Gallagher, "Using Vanishing Points To Correct Camera Rotation In Images," in *Proceedings of the 2nd Canadian conference on Computer and Robot Vision*: IEEE Computer Society, 2005.

[10] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 22, pp. 1330-1334, 2000.

[11] L. Dawei, H. Qingming, L. Yang, Z. Guangyu, and G. Wen, "Video2Cartoon: A System for Converting Broadcast Soccer Video into 3D Cartoon Animation," *Consumer Electronics, IEEE Transactions on,* vol. 53, pp. 1138-1146, 2007.