Automatic Camera Calibration Technique and its Application in Virtual Advertisement Insertion System

Shijin Li, Bing Lu
School of Computer & Information Engineering
Hohai University
Nanjing, China, 210098
lishijin@hhu.edu.cn

***Abstract-* This paper has proposed a novel approach to automatic camera calibration for the virtual advertisement insertion system. The presented technique can extract feature lines in the field automatically, through which the internal and exterior parameters of camera can be obtained. Firstly, the adaptive segmentation of color image is carried out to the football field images. Secondly, the feature lines are extracted by applying a multiple Hough transforms to the segmented images. The feature points in the image space which are corresponding to the reference coordinate of world are generated automatically from the intersections of the virtual grid-lines, which are used for automatic camera calibration. Then the advertisements can be inserted into the soccer field in the appointed positions according to these parameters. Experimental results have shown that this method can meet the needs of the virtual advertisement insertion system and be put into use in real application scenarios.**

I. INTRODUCTION

Advancements in image processing, computer vision, and computer graphics have made real time application of virtual or augmented reality (VR/AR) in sports programming feasible. Virtual imaging technology is being used to augment television broadcasts - virtual objects are seamlessly inserted into the video stream to appear as real entities to TV audiences. Virtual advertisements, the main application of this technology, are providing opportunities to improve the commercial value of television programming while enhancing the contents and the entertainment aspect of these programs [1]. A typical virtual advertisement insertion system consists of three subsystems, i.e., camera tracking subsystem, real-time virtual ad generation subsystem and video composition subsystem. Camera calibration is the pre-requisites for the running of camera tracking subsystem. Figure 1 gives the architecture of our virtual ad insertion system, which is developed by Hohai University in China.

There are many existing camera calibration algorithms in the literature [2], such as non-linear optimization methods [3], linear techniques which compute the transformation matrix [4] and two step techniques [5]. The two-step techniques use a linear optimization to compute some of the parameters and, as a second step, the rest of the parameters are computed iteratively. These techniques permit a rapid calibration considerably reducing the number of iterations. Moreover, the convergence is nearly guaranteed due to the linear guess obtained in the first step. In our implemented system, Tsai's method [5] is adopted, for its rapidness and stability. However, there are some limitations of that method, such as the number of feature points should not less than five (the soccer field can be thought as co-planar) and should be distributed broadly across the field of view, which is difficult to ensure in practice.

In this paper, we put forward a novel method to fulfill this request, which generates many virtual feature points as the input of Tsai's method. Since there are many field lines in the soccer match image, we extract and recognize four lines of the penalty area and compute the intersections of grid lines between the four lines by the homography between the four lines in the image and their counterpart in physical court. Compared with manual input of the intersections of the filed lines, our method is more accurate and stable, and can be put into use in real application scenario.

The rest of the paper is organized as follows. Section 2 introduces the extraction of the field lines from soccer match images. Section 3 gives the proposed calibration method. How to utilize the parameters to insert the virtual advertisements is described in section 4. In section 5, the experimental results are analyzed. Finally, conclusion is drawn and some future work is also pointed out.

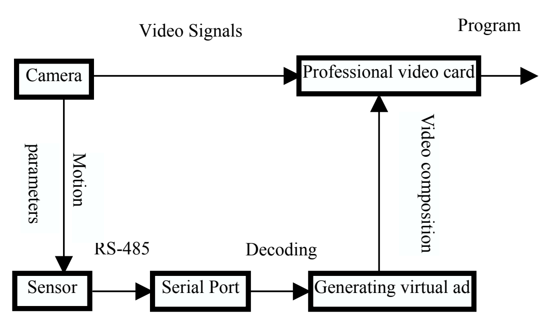

Fig 1. The architecture of our virtual advertisement insertion system.

II. THE SEGMENTATION AND EXTRACTION OF THE FIELD LINES

A. The segmentation

The field lines which are usually made up of white points are on the green grass. The football playground must be

1-4244-0737-0/07/$20.00 ©2007 IEEE

segmented firstly so that the interference of the audiences and the lines which are out of the football field can be ignored in our work. The HSI color space is chosen and it is supposed that the football field is the maximal connected region in the image which means that the peak of the histogram represents the dominant color of the football field. The color football field image is segmented with 3σ rule according to the peak and computed covariance of the histogram. Figure.2 shows the histogram of H component of a soccer match image. Figure 3 shows the segmented soccer field which is represented in white points.

The Laplacian of Gaussian filter(LoG) is used to segment the field lines, because it can highlight the field lines on the grassfield. We use the LoG operator to convolve with the gray image. After convolution, if the result value is lager than a specified threshold, then the point is recognized as on the field lines. All the points which are lager than the threshold are set to white, while others to black. Then we can get a binary image of field lines. Some post-processing must be conducted to the binary image, such as the morphological operators. It can remove the noise points which are produced by the players. The number of the field points in the Hough transform will be reduced so that the computation time of the Hough transform will be reduced too. Figure.4 gives the result of the segmentation of the field lines. Figure.5 gives the resulting image in which the noises are removed by morphological operations.

B. Hough transform

Hough transform is one of the classical methods for detecting straight lines and curves in the digital image. The main idea of it is the dual problem of point and line. All points on the same line in the image space will intersect at one point in the HT parameter space. All lines intersected at one point in the HT parameter space are corresponding to a line in the image space. Hough transform reformulates the line detection problem from the image space to the HT parameter space, which can be carried out by the accumulation in the HT parameter space.

In the conventional method, a peak in the HT parameter space will be found firstly. Then the cells around the peak is reset to zero. All the field lines on the soccer court can be divided into two groups and they are parallel in the physical football field. Two problems will we meet if the conventional method is utilized [6]. The first one is that we can't clear up all the cells around the peak with large values if the window is too small, which will lead to the unsatisfying situation that we will find the same field line in the next round Hough transform. The second problem is that we will clear up the peak value of short lines if the window is too large. Both the two groups are close to each other in angles and the difference of their intercepts are small, so we can't extract all the field lines and it will make some trouble in the following recognition of the model of the football field. We hope to find all the field lines without errors, especially the short ones.

We apply multiple Hough transform to the segmented images to extract the field lines. After one field line is extracted, we will erase it in the image space. The slope and the intercept of the line can be computed after the location of the peak. Then the position of the field line in the image space is known. In order to reduce the influence of the width of the lines, we adopt the least square method to fit the curve to correct the parameters of the extracted lines, thus improving the accuracy. After erasing the "inflated"line in the HT space, it will not find out the line again in the next round Hough transform. So the extraction of all the field lines, including the short ones, becomes easy.

C. The recognition of the field line

Szenberg et al. [7] have done some research on automatic camera calibration for image sequences of a football match. They utilize covariance matrix to extract the points on the feature lines firstly. Then they use a model-based recognition method in the recognition of the feature lines. A set of restrictions are given first, then an interpretation tree structure is used to search the right model. In contrast, in our method, only 4 lines should be recognized, thus largely ease the complexity of the model.

According to the structure of the soccer field, the recognition of the lines can be divided into two groups: the left half-court and the right half-court. In the following, the right half-court case is illustrated. Firstly, all the extracted lines of the soccer field can be divided into two groups and they are parallel in the physical football field. Secondly, the slope of the lines which are parallel with the goal line is negative, while the other group is positive. The candidate field lines can be divided into two group based on their slopes. The group whose slope is less than zero can distinguish the goal line and the other two lines by the line's intercept because the goal line has the maximal intercept. We can recognize the other group by the intersection information with the recognized two lines. It is easy to recognize the penalty area lines with the intercept information.

III. AUTOMATIC CAMERA CALIBRATION

The feature points in the image which are needed by the automatic camera calibration are produced by the virtual gird-lines. The world points and the image points can be paired by automatic camera calibration. Then we can insert the advertisements to the appointed positions in the soccer field. So automatic camera calibration is the key step of the virtual advertisements insertion system.

A. The generation of the virtual gridding points

The feature points in the image space which are corresponding to the reference coordinate of world are produced automatically from the intersections of the virtual grid-lines. According to the principle of the least square method, we know that a good calibration result can be got when there are many feature points and they are distributed uniformly in the image. Twenty feature points are used in this paper for automatic camera calibration. The penalty area can be divided by four penalty area lines and five extra virtual gird-lines. In the direction of the goal line, we partitioned the penalty area into three parts. In the vertical direction, we partitioned the penalty area into four parts. Twenty feature points can be got by resolving the intersections of the nine lines. The physical coordinates of the twenty feature points are calculated according to the model of the soccer field. Thus the two group of feature points are used as input to the automatic camera calibration algorithm.

Five virtual grid-lines can be got by the homography of the two planes: the physical plane and the one in the image. Homography gives out the relationship of two images from different viewpoints [8]. Assume that X and X' are two projection points of one point on the plane π . It fits the following equation:

$$X' = HX \quad (1)$$

More specifically, let (x, y) and (x', y') are two projection points of one point from the left and right viewpoints. Then formula (1) can be re-written as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

The equalization is with some scale constant. Here (x, y) and (x', y') denotes image points and physical points respectively. Eight equations can be got from four intersection points in the image and their counterparts in the physical soccer field. We set h_{33} in the formula (2) to 1. Then the omography of the plane can be got by resolving these equations. (We can use the least square method to get Homography when the number of feature points is more than 4)

We can map the penalty area of the football image to a physical model which have already been grided. Then the penalty area can be grided by the resolved homography, and the virtual grid-points are produced. In the experiment, twenty feature points are used.

B. Camera calibration

The algorithm of Tsai's two-steps camera calibration based on the radical alignment constraints [5] is used in our system. Each feature point pair can make an equation. The number of the unknown variables in the joint equations is smaller than the number of the equation. It is over-determined. The intrinsic and extrinsic parameters of camera are got by the least square method. This method is simple, quick and accurate. The method of Tsai assumes that there are some parameters of the camera which are provided by manufacturers. This fact reduces the number of calibrating parameters in the first step where an initial guess is estimated. Moreover, although all the parameters are iteratively optimized in the last step, the number of iterations is considerably reduced.

IV. VIRTUAL ADVERTISEMENTS INSERTION

The intrinsic and extrinsic parameters of the camera are obtained after the automatic camera calibration. These parameters can map a physical point into an image point. The positions of the virtual advertisements in the image can be computed if its positions in the physical coordinates are known. We must make sure the positions, the shape and the size after the transformation of perspective projection in the reference coordinate of world, such as in one or two sides of the goal. The four vertexes of the advertisements in the image space can be computed after camera calibration, through which we can get the positions of the virtual advertisements. Then the whole advertisements image will undergo the same processing. After that, the advertisement images are inserted into the stream of the background video. Finally the composite video signal is broadcasted into the air.

V. EXPERIMENTAL RESULTS

We have conducted the following experiments to compare the accuracy of the proposed method and the manual input manner.

In order to validate the accuracy of the method, we have done ten times calibrations with manual input feature points and one time automatic camera calibration for a given image.

The position of the virtual advertisements can be computed by the parameters of camera after calibration. All the virtual advertisements are nearby the goal line and one vertex of the virtual advertisement overlaps one of the goalposts. To ease the comparison, we only compute the four vertices positions of the inserted virtual advertisement. There are five times the positions of the advertisement deviating inside the playground, three times the positions of the advertisement deviating outside the playground, and only twice the positions of the advertisement are on the goal line. From above we can see that the performance of manual calibration is not stable. The automatic camera calibration result is much better. Table I gives the detailed results of the experiments.

Form Table I we can see that the results of manual camera calibration are diverse. The maximal error value of the vertex is 13 pixels. So it affects the positions of the virtual advertisement and the result of calibration is instable. It's mainly influenced by the manual input of the coordinates. When we input the points, it's difficult to observe the discrepancy of 1-2 pixel(s) by eyes. Accordingly, it affects the precision of the calibration results later. In addition, the available feature points are comparatively few, so we can choose 4-5 points in the same line. It's difficult to ensure whether the points are in the same line or not, so it also affects the precision of the camera calibration.

In summary, the results of manual calibration are not stable and the proposed automatic method can meet the needs of virtual advertisement insertion system.

Figure 6 shows the generated grid-lines in the penalty area. Figure 8 shows the result image after inserting virtual advertisements. The original advertisement images are given in Figure 7.

It should be noted that the calibration procedure is only executed once when the virtual ad insertion system is launched. So the proposed method can take 1~2 seconds to get the results. Afterwards, the sensors for camera tracking will generate the parameters for real-time rendering of the virtual ad.

TABLE I.
The coordinate of the resulting vertexes by hand and by automatic method

	Vertex 1	Vertex 2	Vertex 3	Vertex 4	Acc
M1	(216,186)	(132,222)	(161, 176)	(76, 211)	**In**
M2	(213, 185)	(130, 221)	(158, 174)	(73, 209)	**In**
M3	(211,184)	(128, 219)	(155, 173)	(71, 206)	**On**
M4	(208, 186)	(123, 221)	(151, 177)	(63, 211)	**Out**
M5	(214, 186)	(131, 221)	(159, 175)	(74, 209)	**Out**
M6	(213, 185)	(130, 220)	(158, 176)	(73, 209)	**In**
M7	(209, 185)	(124, 220)	(151, 176)	(64, 210)	**Out**
M8	(210, 185)	(127, 221)	(154, 176)	(70, 209)	**On**
M9	(215, 186)	(131, 222)	(160, 175)	(76, 210)	**In**
M10	(208, 185)	(123, 222)	(152, 174)	(63, 209)	**Out**
Max	(8, 2)	(9, 3)	(10, 4)	(13, 5)	
Avg	(211, 185)	(128, 221)	(156, 176)	(70, 209)	
A	(210, 186)	(128, 219)	(155, 175)	(71, 208)	**On**
Note: **Mi**-manual input ith time; **A-** automatic input; **Max-** max error of manual input; **Avg**- the average of four vertices position; **Acc**-Accuracy; **In**- inside the line; **Out**- outside the line; **On**- on the line					

VI. CONCLUSIONS

This paper describes a novel approach to automatic camera calibration for the virtual advertisement insertion system. The main contribution lies in that through generating virtual feature points in the image, automatic camera calibration can be accomplished easily, without the user's intervention. We segment the color image firstly, then the field lines are extracted from the segmented image by using multiple Hough transforms. By the computation of the homography between the image plane and the physical plane, many feature points can be generated and thus solves the problem of the lackage of feature points for the camera calibration algorithm. The proposed method has been put into practical use in our virtual advertisement insertion system successfully. Possible future work includes the extension to other sport games, such as tennis match.

REFERENCES

[1] Y. Tan, "Virtual Imaging in Sports Broadcasting: An overview, " in Proceedings of International conference on virtual reality and its application in industry, Hangzhou, China, 2002, pp. 330-339.

[2] J. Salvi, X. Armangue, and J. Batlle. "A comparative review of camera calibrating methods with accuracy evaluation." Pattern Recognition, 2002, 35(7): 1617～1635

[3] C.C. Stama, C. Theurer, and S.W. Henriksen, *Manual of Photogrammetry*, 4th Edition, American Society of Photogrammetry, Falls Church, VA, 1980.

[4] O.D. Faugeras, G. Toscani, "The calibration problem for stereo," Proceedings of the IEEE Computer Vision and Pattern Recognition, 1986, pp. 15－20.

[5] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE Int. J. Robot. Automat.RA-3 (1987), pp. 323－344.

[6] Y. Furukawa，Y. Shinagawa，"Accurate and robust line segment extraction by analyzing distribution around peaks in Hough space. " Computer Vision and Image Understanding 92 (2003): 1－25

[7] F. Szenberg et al, "Automatic Camera Calibration for Image Sequences of a Football Match. " in Proceedings of ICAPR 2001, LNCS 2013, pp. 301-310.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, Cambridge, UK, 2000.

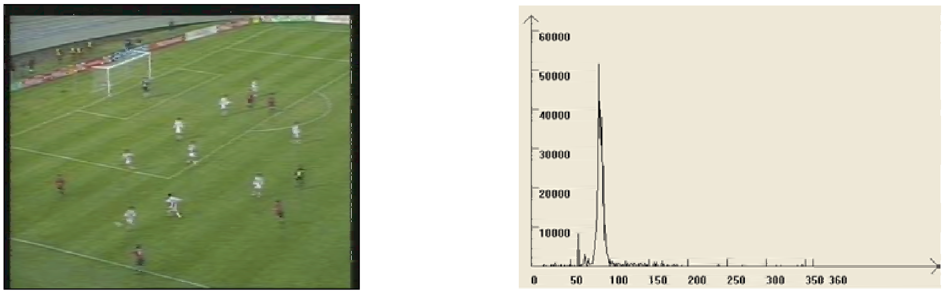

Fig 2. A frame of image from one soccer match and the histogram of its hue component.

Fig 3. The segmented field area

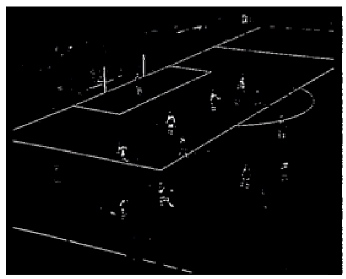

Fig 4. The field line and noise

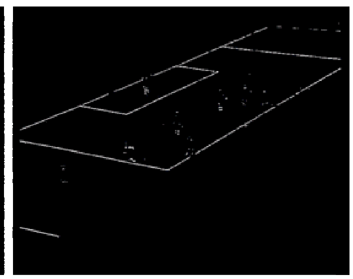

Fig 5. after post-processing

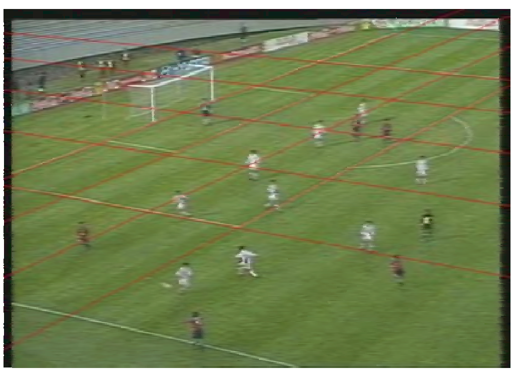

Fig 6. The generated grid-lines in the penalty area.

Fig 7. The original images of two ads.

Fig 8. The resulting images after insertion of virtual advertisements.