Position Estimation of Solid Balls from Handy Camera for Pool Supporting System

Hideaki Uchiyama and Hideo Saito

Keio University, 3-14-1 Hiyoshi, Kohoku-ku 223-8522, Japan

Abstract. This paper presents a method for estimating positions of solid balls from images which are captured using a handy camera moving around the pool table. Since the camera moves around by hand in this method, the motion of the camera in 3D space should be estimated. For the camera motion estimation, a homography is calculated by extracting the green felt region of the table-top area that is approximated to a polygon. Then, the balls are extracted from the table-top region for obtaining the positions of the balls. The 3D position of each ball is estimated using a projection matrix determined by the homography. The ball areas are classified by distribution of RGB data in each area. We apply our method to image sequences taken with a handy camera for evaluating the accuracy of the ball position estimation. By this experiment, we confirm that the accuracy of the estimated position is up to 18mm error, which is sufficiently small for displaying the strategy information in the pool supporting system.

1 Introduction

Pool is one of complex sports in the world that need knowledge of physical laws to decide direction and strength of a shot based on the arrangement of balls. However, it is difficult for beginners to shoot a ball considering dynamic behavior because they concentrate on shooting the center of the ball accurately. Since they can shoot the ball easily if the system teaches users the desired tracks of the ball, some pool supporting systems are proposed. Jebara used a head mounted live video display (HMD) with a camera and lines of a desired trajectory of the balls are rendered on HMD[1]. Also, Larsen used a computer controlled laser pointer to show a correct layout of the balls and the target[2][3][4]. However, such equipments are not desirable in pool game because playing pool with HMD is not natural for users, and installing the laser pointer system in a usual environment of pool hall is difficult. To be used popularly, the system needs to be composed of small equipments such as a cellular phone that have a camera, a small screen and CPU.

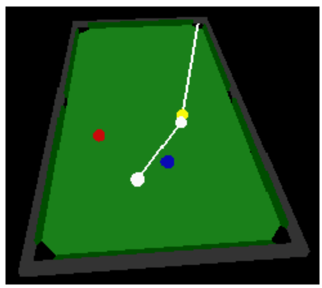
This paper describes a method for estimating the positions of solid balls from image sequence taken with a handy camera. The estimated positions are then used for showing supporting information on a LCD display. As for the balls, there are stripe balls and solid balls in pool game. In Chua's research, they classify two kinds of balls for Eight-Ball game which is one of the most famous pool games in the world[5]. Since we intend to apply our method to Nine-Ball game which is also a famous pool game, we propose the method for classification of colors of solid balls.

L.-W. Chang, W.-N. Lie, and R. Chiang (Eds.): PSIVT 2006, LNCS 4319, pp. 393–402, 2006.

A Ball become a circle and a table becomes a rectangle as a 2D objects when an image is captured from table top[5][6][7][8]. However, we treat balls and a table as a 3D object since it is difficult for users to capture images from the table-top. In our method, we assume that images as shown in Fig.1 (a) are captured with a handy camera by moving around the pool table. To each image, we extract a green felt area to compute a homograpy by using planarity of the area. After extracting the area, we extract circular regions as balls and then determine the center positions of the circles. 3D coordinates of the centers of balls can be estimated using a projection matrix computed by the homograpy of the table-top plane. To avoid accidental miss-detection of the balls, we use multiple frames in the input image sequence. Then users judge the correctness of the estimated 3D positions by watching the arrangement of the balls.

We apply our method to image sequences taken with a handy camera for evaluating the accuracy of the estimated ball positions. By this experiment, we confirm that the accuracy of the estimated position is up to 18mm error, which is sufficiently small for the pool supporting system.

(a) Input



(b) Desired output

Fig. 1. Input and output

2 Proposed Method

In Jebara's research[1], captured images from eye position were divided into a green felt region, balls, corners, and pockets by using a probabilistic color model for understanding the status of the table-top. The white lines which connected a target ball with a desired pocket to provide strategic information to players. Because they don't compute the geometrical relationships of all the balls and others, they can display only simple lines. In our research, we compute the 3D positions of all the balls on the table by capturing images of the whole parts of the table for pool supporting system based on the arrangement of all the balls.

2.1 Overview

Input images are of the whole of the pool table captured from different viewpoints, and then our method is applied to each image.

First, a green felt region of the table is extracted by color segmentation to compute a homography. Then, the region is approximated to a quadrangle because four vertices are necessary for computing the homography. After computing the homography, a projection matrix is computed from the homography by Simon's method[9][10][11].

Next, non-green areas in the green felt area are extracted as candidates of balls. Pockets and cushions are included in the candidates. Balls are extracted by the size of an area and the degree of circularity in each candidate. For each ball, the number of the ball is determined by distribution of RGB data.

In the end, 3D positions of the balls are computed by using the projection matrix and the 2D coordinates of the centers of the balls in images. To decrease false detections, a user determines whether the result is sufficient or not.

In our research, we implement the proposed algorithm by using OPENCV[12], and introduce the functions which we use.

2.2 Green Felt Area Detection

It is necessary for computing a homography to detect a plane. Both a frame and the inside of the pool table which are made of a green felt are planes shown in Fig. 2(a). Because of cushions, the whole part of the inside cannot be detected like the parts of the red circle of Fig. 2(a). All regions of the green felt that are a frame and the inside are extracted in our method.

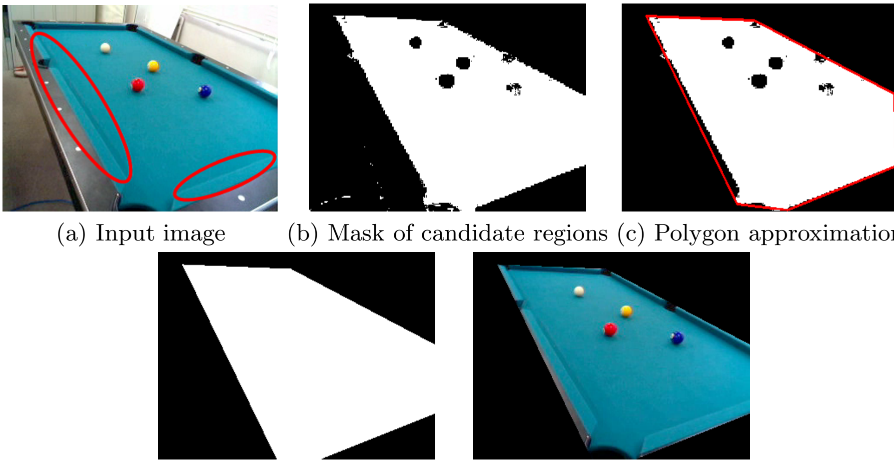
To extract the green felt area by color segmentation, RGB vector of the area is measured beforehand as a template. By computing the angle between RGB vector of each pixel and that of the template, each pixel is determined whether included in the area or not. Each pixel is included in the area when the angle is below a threshold because the angle is small if the colors of two vectors are close. Fig. 2(b) depicts the mask of candidates of the area.

Next, the largest region in the mask is extracted, and approximated to a polygon as shown in Fig. 2(c) using Approxpoly() of the OPENCV function. From the sides of the polygon, four sides from longer ones are chosen, and the quadrangle is made by these sides as the mask of the green felt area in Fig. 2(d). Fig. 2(e) shows the green felt area which is extracted from a input image segmented by the mask.

2.3 Camera Calibration

The green felt area is extracted and approximated to a quadrangle in Section. 2.2 to compute a homography $\boldsymbol{H}$. The homography is computed by associating the vertices of the green felt area of an actual pool table and those of the quadrangle in a input image. To compute a projection matrix by the homography, we employ Simon's method.

A 3D coordinate system is related to a 2D coordinate system by 3×4 projection matrix $\boldsymbol{P}$. Thus, each 3D coordinate system designed for each plane is also related to the input images by each projection matrix. If a Z coordinate of each plane is set to 0, the homography $\boldsymbol{H}$ also relates between each plane and the input images. In Eq. (2), the projection matrix $\boldsymbol{P}$ is composed of intrinsic



(a) Input image (b) Mask of candidate regions (c) Polygon approximation

(d) Mask of a green felt area (e) Extracted green felt area

Fig. 2. Green felt area detection

parameters $\boldsymbol{A}$, a rotation $\boldsymbol{R}$ and a translation $\boldsymbol{t}$ of extrinsic parameters. Also, the homography $\boldsymbol{H}$ is expressed in Eq. (3) which is the deleted $\boldsymbol{r}_3$ of $\boldsymbol{P}$.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \simeq \boldsymbol{P} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \simeq \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \simeq \boldsymbol{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{1}$$

$$\boldsymbol{P} = \boldsymbol{A}\left[\boldsymbol{R} \mid \boldsymbol{t}\right] = \boldsymbol{A}\left[\boldsymbol{r}_1 \boldsymbol{r}_2 \boldsymbol{r}_3 \boldsymbol{t}\right] \tag{2}$$

$$\boldsymbol{H} = \boldsymbol{A}\left[\boldsymbol{r}_1 \boldsymbol{r}_2 \boldsymbol{t}\right] = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{3}$$

First, intrinsic parameters $\boldsymbol{A}$ are computed by the homograpy $\boldsymbol{H}$. We assume that the skew is 0, the aspect ratio is 1 and the principal point is the center of the image. The intrinsic parameters can be defined as in Eq. (4). Then, we only have to estimate the focal length f. Using the property of the rotation matrix $\boldsymbol{R}$ that is the inner product of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ is equal to 0, the focal length f is computed as shown in Eq. (5).

$$\boldsymbol{A} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} (c_x, c_y) & : & \text{principal point} \\ f & : & \text{focal length} \end{pmatrix} \tag{4}$$

$$f = \frac{(h_{11} - c_x h_{31})(h_{12} - c_x h_{32}) + (h_{21} - c_y h_{31})(h_{22} - c_y h_{32})}{-h_{31} h_{32}} \tag{5}$$

Next, the rotation $\boldsymbol{R}$ and the translation $\boldsymbol{t}$ of extrinsic parameters are computed. $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{t}$ are computed by the homography $\boldsymbol{H}$ and a inverse matrix of intrinsic parameters $\boldsymbol{A}$ shown in Eq. (6). $\boldsymbol{r}_3$ is computed by the nature of the rotatin matrix that the cross product of $\boldsymbol{r}_1$ and $\boldsymbol{r}_r$ becomes $\boldsymbol{r}_3$ in Eq. 7.

$$[\boldsymbol{r}_1\boldsymbol{r}_2\boldsymbol{r}_3\boldsymbol{t}] = \boldsymbol{A}^{-1}\boldsymbol{H} \tag{6}$$

$$\boldsymbol{R} = \left[\boldsymbol{r}_1\ \boldsymbol{r}_2\ (\boldsymbol{r}_1 \times \boldsymbol{r}_2)\right] \tag{7}$$

2.4 Ball Detection

Balls are detected in the green felt area because the balls surely exist on the area. To detect the balls, non-green areas are extracted in the green felt area as candidates of the balls, and then balls are detected from candidates.

To extract the candidates of balls by color segmentation, RGB vector of the green felt used in Section. 2.2 is used as a template. For each pixel, the angle between RGB vector of each pixel and that of the template is computed. Also, the difference of Norm of two vectors that shows the difference of the brightness is calculated to determine whether the green felt is or not because there is the similar color of the felt in the colors of the balls. Each pixel is determined as a candidate of balls when the difference of the norm is over a threshold and the angle of two vectors is over another threshold. Fig. 3(b) shows candidates of the balls. Each candidate is determined whether it is the pixel of the balls or not by the size and the degree of circularity. The mask of the balls is made as shown in Fig. 3(d). Fig 3(e) depicts the balls which segmented by the mask from a input image.

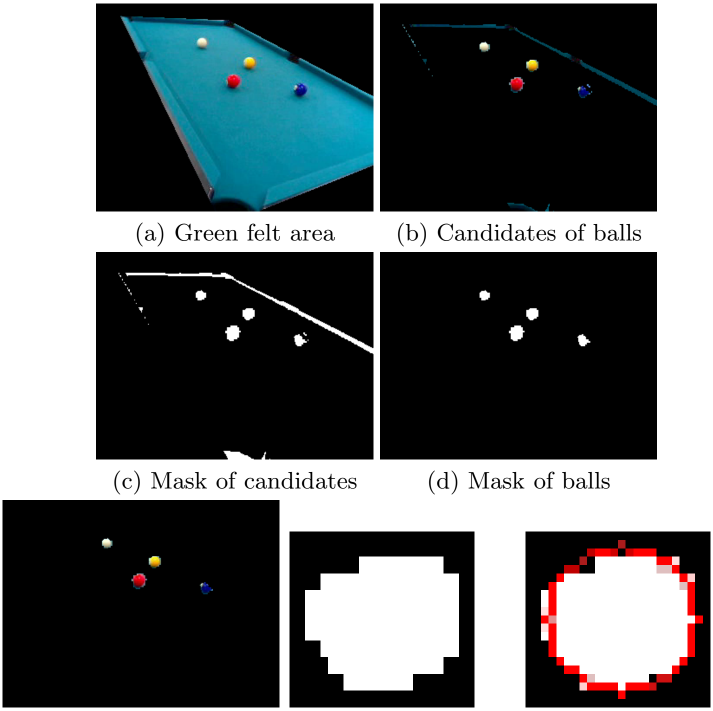
A ball in 3D world becomes a circle in 2D image. Then, each area of the balls in Fig. 3 is approximated by the circle in Fig. 3(f) by using cvMinEnclosingCircle() of the OPENCV function and the center coordinate of the circle is also calculated by the function to compute the 3D coordinates of the ball.

2.5 Classification of Solid Balls

We propose the classification of solid balls which colors are nine.

In the ball area depicted in Fig. 4, the shadow of the ball and the specular reflection element are also included. If the average of the color in the area is computed to classify the ball, miss-classification often occur because of the influence of the shadow and the specular reflection element strongly. To classify the balls precisely, we apply a voting such that the closest color of the balls is selected.

RGB vector of each ball is measured beforehand as a template. The angle between RGB vector of each pixel in the ball area and that of the template of each ball is calculated, and votes on the balls to which the angle is minimum like Table. 2.5 are done. Number zero is a white cue ball. Then, the ball with lion's share of votes is selected.



(a) Green felt area (b) Candidates of balls

(c) Mask of candidates (d) Mask of balls

(e) Balls (f) Mask of a ball (g) Circle approximation

Fig. 3. Detection of balls

2.6 Three Dimensional Coordinate of the Balls

In Eq. (8), (x, y) is 2D coordinates in the images and (X, Y, Z) is 3D coordinates in the world coordinate system. The projection matrix $\boldsymbol{P}$ is calculated by the homography in Section. 2.3.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \simeq \boldsymbol{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{8}$$

It is necessary to compute (X, Y, Z) to determine 3D coordinates of the center of the ball. However, it is not possible to compute from (x, y) in the image and the projection matrix $\boldsymbol{P}$ because the degree of freedom of 3D coordinates is three. Then, Z is calculated beforehand by nature of a sphere to decrease the degree of freedom.

As Fig. 5 is shown, the straight line that connects the point p in the image and the point$P$ in the world which appear in the image passes the center$C$ of the ball by nature of a sphere. X and Y can be computed because the degree of freedom becomes two if Z is known. Since the plane which used when the

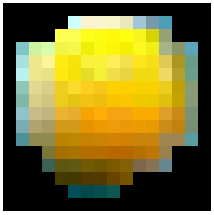


Fig. 4. Distribution of color

Table 1. Judgement of balls by vote

ball number	0	1	2	3	4	5	6	7	8	9
result of vote	23	83	0	0	0	0	32	0	15	0

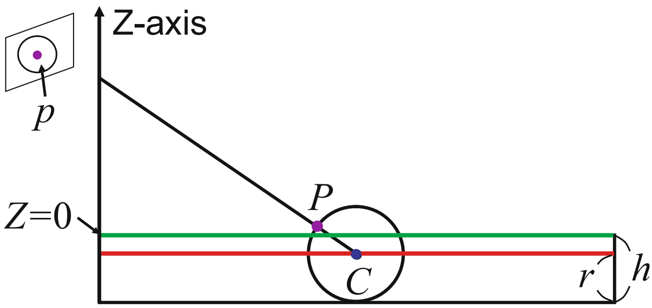


Fig. 5. Computation of the three dimentional position of a ball center

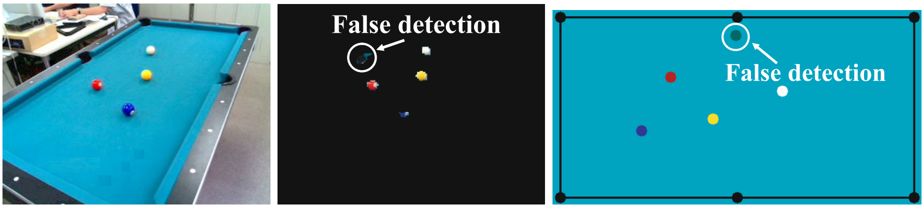
homography is calculated is $Z = 0$, Z of the center of the balls is determined as $Z = -(h-r)$ because the radius r of the ball and the height h of the cushion of a pool table are known. X and Y can be computed by using $\boldsymbol{P}$, (x, y), and Z.

2.7 Judgment of the Result by User

When the 3D positions of the balls are estimated in each image, the pocket or parts of the pool table might be mistakenly detected as a ball in Fig. 6(b). To decrease such false detections, a user determines whether the estimated positions of the balls are correct or not by watching the display of the arrangement of the balls from the top of the table as shown in Fig. 6(c) when the positions can be computed for the first time. This is repeated until the user satisfies the estimated arrangement of the balls. As for the estimated positions in the image taken from other aspects, correctness is automatically judged depending on the distance with the position of the ball that has already been computed. Then, the average of the computed positions is calculated. In the way, the false detections can be removed by user interaction.

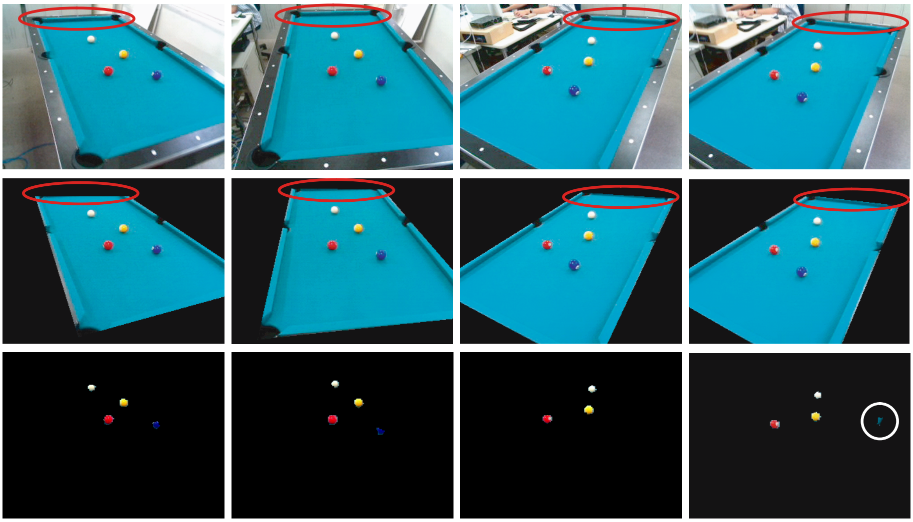
3 Experiments

We apply our method to some image sequences including 150 images taken with a handy camera for evaluating the accuracy of the estimated 3D ball position



(a) Input (b) False detection (c) Display for users

Fig. 6. Judgment of the result by user



(a) frame 20 (b) frame 50 (c) frame 80 (d) frame 90

Fig. 7. Results

estimation. As Z coodinate is constant depicted in Section. 2.6, we confirm the accuracy of (X, Y).

The size of table is 1330mm×700mm and that of images is 320×240pixels. We used four balls which colors are white, yellow, blue and red. We put on the balls which coordinate is measured with a tape measure beforehand, and then we compare their coordinates with the coordinates estimated from the input images. Fig. 7 show results that the first row shows inputs, the second row shows grenn felt areas, and the third row shows balls.

First, we count the number that balls are extracted from 140 images. The result is shown in Table. 2. Since we captured images freehand, 34 images blurred and all part of a board was not include in these images . We couldn't extract a board from such images, and alse balls. White, yellow and red balls can be extracted in many images because their colors are quite different from the color of the green felt. On the other hand, a blue ball cannot sometimes be extracted

Table 2. Number of balls extracted

white	yellow	blue	red	corner
103	116	82	116	5

because its color is a little close to the color of green felt. Pocket is also extracted as a ball shown in a ball area image of Fig. 7(d) because the shape of pocket is close to the circle and the color is different from that of the green felt.

Next, we compare the ball positions measured with a tape measure with the estimated positions from the input images via the proposed method. Table. 3 shows the result of comparison. The result shows that Y values contain more errors than Xvalue. The error is mostly caused by the unstable detection of the far side edge of the green felt region. As shown in Fig. 7, the far side edge indicated by the red ellipse cannot clearly be captured in the input image. In Fig. 8, square dots show ground truth positions, while circular dots show the estimated positions from the input images. Even though there are around 18mm errors in Y components, the arrangement of the balls estimated via the proposed method is similar to the ground truth, therefore we consider that the proposed method can sufficiently be used for the pool supporting system.

Next, the center position of the white ball computed in each frame are shown in Fig. 9. The error in each frame is not constant because the lighting condition is different, depending on the direction where a pool table is captured.

Table 3. Ball positions(m)

ball color	white		yellow		blue		red	
coordinate	x	y	x	y	x	y	x	y
ground truth	0.373	0.823	0.273	0.573	0.223	0.323	0.423	0.423
estimated from images	0.378	0.841	0.274	0.582	0.228	0.317	0.429	0.425

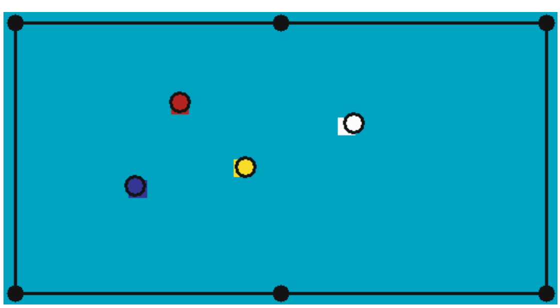


Fig. 8. Comparison of ground truth (rectangular dots) with estimated from image (circular dots)

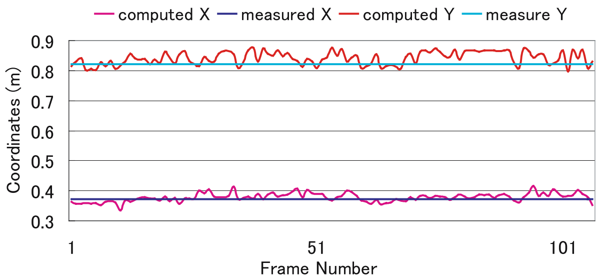


Fig. 9. Coordinates of white ball

4 Conclusions

We have proposed a method for estimating positions of solid balls from images which are captured using a handy camera moving around the pool table. For estimating the positions of all the balls on the pool table, this method first computes homographies of the table-top region by extracting the green felt area via color segmentation. The computed homographies provide camera parameters so that the 3D positions of the balls can be estimated from the extracted positions of the balls in the input images. In the result, we have shown that the position estimation accuracy is up to around 18 mm error, which is sufficiently accurate for the pool supporting systems.

References

1. T.Jebara, et al.: Stochastics: Augmenting the Billiards Experience with Probabilistic Vision and Wearable computers, Proc. of ISWC, pp. 138-145, 1997.
2. L.B.Larsen, et al.: The Automated Pool Trainer - A multi modal system for learning the game of Pool, Proc. of ICIMADE, pp. 90-96, 2001.
3. L.B.Larsen and T.Brondsted: A Multi Modal Pool Trainer, Proc. of IPNMD, pp. 107-111, 2001.
4. L.B.Larsen, M.D.Jensen and W.K.Vodzi: Multi Modal User Interaction in an Automatic Pool Trainer, 4th IEEE ICMI, pp. 361-366, 2002.
5. S.C.Chua, E.K.Wong and V.C.Koo: Pool Balls Identification and Calibration for a Pool Robot, Int. Conf. on ROVISP, pp. 312-315, 2003.
6. S.C.Chua, E.K.Wong and V.C.Koo: Simulation of A Robotic Pool, M2USIC, 2004.
7. S.C.Chua, et al.: Decision Algorithm for Pool Using Fuzzy System, Proc. of iCAiET, pp. 370-375, 2002.
8. M.E.Alian, et al.: Roboshark: a Gantry Pool Player Robot, 35th ISR, 2004.
9. G.Simon, A.W.Fitzgibbob and A.Zisserman: Markerless tracking using planar structures in the scene, Proc. of the ISAR, pp. 120-128, 2000.
10. G.Simon and M.Berger: Reconstructing while registering: a novel approach for markerless augmented reality, In Proc. of the ISMAR, pp. 285-294, 2002.
11. Y.Uematsu and H.Saito: Vision-based Registration for Augmented Reality with Integration of Arbitrary Multiple Planes, 13th ICIAP, pp. 155-162, 2005.
12. Intel: http://www.intel.com/