A calibration method for performance analysis in table tennis

Sho Tamaki*
Hideo Saito*

A calibration method for performance analysis in table tennis is proposed. Currently, the common method for 3D measurement in sports match scenarios is to compute projection matrix from 3D-2D corresponding points with time-synchronized cameras. There, however, exist several difficulties to conduct it. The proposing method in this paper makes calibration and 3D measurement easier because 3D corresponding points and time-synchronized cameras are not required in this method. It requires only traveling balls in playing area. The key idea of the method is to use 2D ball trajectories to estimate the temporal offset between two cameras by using a traveling ball. It is experimentally demonstrated that the proposing method could calibrate cameras accurately.

Keywords: calibration, unsynchronized cameras, table tennis

1. Introduction

There has recently been growing interest in the analysis that exploits objective data, so-called performance analysis, in sports sciences. In table tennis, for example, Malagoli et al. (3) clarified the difference in stroking course and footwork styles between asian players and european players. Zhang et al. (12) computed the scoring rate and the usage rate in separated three phase of a rally and analyze the characteristics or the competitive ability of players.

Although various kinds of research have been done in performance analysis, there is no research that uses the velocity of a table tennis ball. There is no doubt about the importance of the velocity of a ball to win a rally in table tennis. The difficulties of measuring the 3D position of a ball can be a reason why the velocity has not been used.

In general, measuring the 3D position of objects in sports match requires multiple cameras whose exposure time is synchronized to each other. In that case, we need to calibrate cameras, namely compute geometrical relation among the cameras. It is done by computing projection matrix with control points, the points whose 3D position is known, or computing intrinsic parameters and fundamental matrix.

It is still a difficult problem to calibrate cameras accurately in practical scenario of sports, such as competition venue. When we need accurate results, the control points or corresponding points need to be inside the space where target objects are supposed to be. However, we cannot enter the inside of courts in sports competition without approval of the competition committee and it is not guaranteed to get the approval. In addition, the exposure time of two cameras have to be synchronized, if the corresponding points supposed to move when we calibrate cameras. Those conditions make the calibration difficult in practical scenario in sports.

In this paper, a calibration method that requires only a traveling ball in a table tennis match is proposed. This method can detect a ball in a match precisely and estimate temporal offset between two cameras and calibrate camera accurately with detected ball positions. Although the accuracy would not overcome the conventional method, which uses time synchronized cameras and control points, there is an big advantage that it is possible to conduct in many scenarios. If the accuracy is acceptable, 3D measurement for performance analysis in table tennis would be conducted easily by using this method.

This paper consists of six sections. Section 2 briefly describes the playing conditions of table tennis and defines the problem of this research. Section 3 introduces related work. Section 4 describes a proposing calibration method. Section 5 describes a experiment that demonstrates that the proposing method could calibrate cameras accurately. Finally, we conclude the paper in section 6.

2. Problem statement

Suppose we calibrate camera without setting arbitrary points whose 3D coordinate is known in playing spaces because it is often unavailable as mentioned in previous section. Furthermore, we do not synchronize the exposure of cameras because our purpose is to make measurement easier. The temporal offset between cameras is also unavailable. Although ball bounces may be the clue for it, to detect and estimate accurate bounce time is difficult problem in low sampling rate, such as 60 fps. It is reasonable assumption that we know the intrinsic parameters of cameras by existing methods, such as the method proposed by Zhang (11), because the intrinsic parameters can be computed out of playing spaces and

* Graduate School of Science and Technology, Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan,
{tamaki, saito}@hvrl.ics.keio.ac.jp

have nothing to do with camera positions.

There are only three objects exist in playing spaces; a table, two or four players, and a ball. The object whose size is known is only the playing surface of a table, 1.525 m × 2.74 m. The length of the other parts of the table are not ruled. We need to compute geometrical relation between two cameras by using these three objects and that is the problem of this research.

3. Related work

We computed the homography between a table and image planes and obtained the rotation and translation of the cameras by decomposing the computed homography [7]. This method, decomposing homography, has advantages that it is applicable in many kinds of case. However, it is difficult to achieve accurate calibration because each camera is calibrated independently with table. The accuracy of triangulation is depended on the accuracy of the estimated rotation and translation, and the separate calibration may accumulate the error of the two times calibration. In addition, the number of points for calibration, four in this case, is too small for enhancing its accuracy. According to Hertly and Zisserman [6], we should collect points as many as possible in practical situation because it is difficult to get image points without noises.

Noguchi and Kato [4] calibrate cameras geometrically and temporally with a moving marker. Their method estimates temporal offset between two cameras and calibrate cameras accurately even if cameras are not time-synchronized. If a marker can be replaced by a table tennis ball, this method would be useful for 3D measurement in table tennis matches. Their method was experimentally demonstrated to be able to conduct accurate calibration. In their method, there are, however, inappropriate assumptions for our case. First, they assume that the motion of the marker is slow or linearly. They interpolate the motion of a ball linearly between successive two frames because of the assumption. A table tennis ball often travels fast and its trajectory may not be linear. Second, they implicitly assume that they can detect markers perfectly and do not handle with false positives. It is, however, difficult to assume that a table tennis ball in a match can be detected perfectly.

4. Proposing method

A calibration method that requires only a traveling ball in a table tennis match is described in this section. Although the exposure time of the cameras are not synchronized and there are not any actual corresponding points, they are estimated accurately in the following procedure.

(1) Estimate rough temporal offset
(2) Detect balls in images
(3) Reconstruct ball trajectories in image space
(4) Update temporal offset and fundamental matrix simultaneously
(5) Compute the extrinsic parameters of the cameras
(6) Compute the scale factor of the extrinsic parameters of the cameras

(1) is done by frame by frame playback manually. It is assumed to include a few frames error. (4) is simple modified version of the method proposed by Noguchi and Kato [4] for dealing with outliers. (2), (3) and (4) are the proposed method and described in the following subsections.

4.1 Detect balls in images The ball detection is done with the color and the shape feature of the table tennis ball, and the feature of its trajectory. The each parameters about a ball and the notation of them are defined as followings;

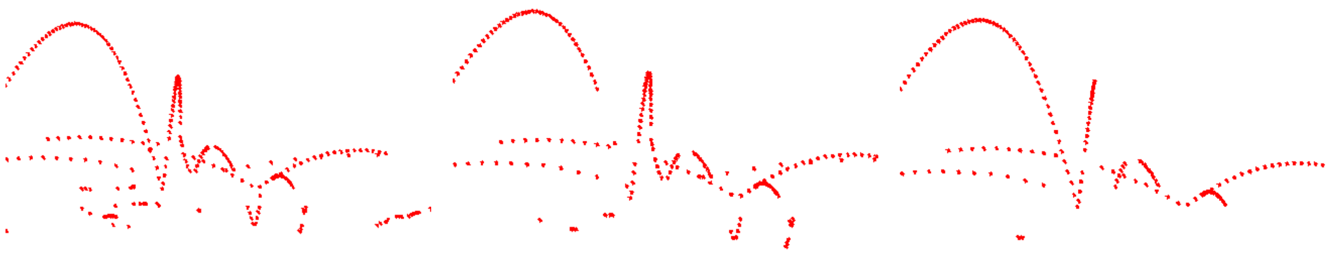
- The upper bound of radius[pixel]: U_r
- The lower bound of radius[pixel]: L_r
- The lower bound of intensity: L_i
- The lower bound of circularity: L_c
- The upper bound of velocity[pixel/s]: U_v
- The lower bound of velocity[pixel/s]: L_v
- The upper bound of the angle between successive two displacement: U_t

U_r, L_r, U_v, and L_v can be determine according to the ratio between actual length and pixel length. U_t can be determined the motion of the imaged balls, which is depended on the position of the cameras. L_i and L_c are depended on the illumination condition or the settings of cameras, such as exposure time. They can be set by obtaining them from captured video before ball detection. The algorithm of ball detection is described in the following part of this subsection.

First, foreground, namely players and a ball, are extracted through background subtraction. We can use the method proposed by Zivkovic [13], which is implemented in OpenCV as backgroundSubtractorMOG2, for background subtraction. Next, the foreground image is segmented based on the contours by using findContours, which is a function in OpenCV. Every segment whose area is greater than U_a is removed because some segments in the players' segments can be false positives. The segments whose intensity is greater than L_i are then extracted because the color of a table tennis ball is nearly white. The boundary length l and the area s of the extracted segments are computed then. The circularity of the segment is defined as following equation and it is computed.

$$C = 4\pi s/l^2. \quad (1)$$

If the circularity is greater than L_c, the segment is extracted. We here have the ball like segments through the ball detection with color and circularity. However, the segments might not center on the center of the ball because the shadow part of a ball is removed when the extraction based on intensity [5]. In order to deal with this problem, the square search range centered on the center of the segment is put on the foreground, which is the image before white area extraction, and the new segment is extracted in the range. Finally, if the area is less than U_a and greater than L_a and the circularity is greater than L_c, the segment is stored as a candidate of the ball. A sample of the detection results is shown in



(a) (b) (c)

Fig. 1. A sample of detection results. (1) shows a result of the detection based on the color and the contour feature. (2) shows a result of the false positive elimination by setting the upper bound about the distance among the consecutive three balls. (3) shows a result of the false positive elimination by setting the upper and lower bound about the distance and the upper bound about the angle among the consecutive five balls.

Fig. 1(a). Although a ball can be detected, false positives are also included in the results. Some false positives are the small segments in player segments, and the others are the background objects that was extracted as foreground because of illumination variation. It is difficult to differentiate an actual ball from them in only the appearance.

As a next step, false positives are eliminated. Let us denote a ball position in ith frame by $\boldsymbol{p}_i$, the function that provides the distance by d, the function that provides the angle by a. If the product of $\mathrm{d}(\boldsymbol{p}_i, \boldsymbol{p}_{i+1})$ and the inverse of the frame rate is less than U_v and greater than L_v, $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i+1}$ are regarded as a ball in consecutive frames. After the consecutive three frames, $\mathrm{a}(\boldsymbol{p}_{i+1} - \boldsymbol{p}_i, \boldsymbol{p}_{i+2} - \boldsymbol{p}_{i+1})$ need to be less than U_t. When the above conditions are satisfied in five consecutive frames, the five candidates are regarded as actual ball. A method that set the upper bound of travel distance among three points has already been proposed [1]. Fig. 1(b) shows a sample of the result of false positive elimination by a method that set the upper bound of travel distance among three points, and Fig. 1(c) shows a sample of the result of false positive elimination by a proposed method. Although the proposed method tend to eliminate many candidates including true positives, it has the advantages in the higher precision, there are very few false positives in the results.

The number of actual balls in a frame is assumed to be is one or zero in this method. This method is not applicable for detecting multiple balls simultaneously due to the assumption. Moreover, if there are some segments that is similar to a table tennis ball in appearance and motion, false positives would be included in the detection results. Those are the limitation of the proposing method. Not only our method, but also the other methods have the same limitation [2] [8] [9] [10]. It is difficult to detect a ball in the images perfectly for now. That's why containing small amount of false positives is assumed in the following sections.

4.2 Reconstruct ball trajectories in image space Ball trajectories are reconstructed as cubic spline curve. The term that trajectories can be reconstructed is limited to the frames where five consecutive balls were detected. The reconstructed trajectories are temporal function, the input of the function is a time and the output is the position. We can estimate ball positions at specific time with the trajectories, even if there is no ball detected at the time. However, the trajectories might provide inaccurate estimation, if the time is out of the range from the time when the first ball was detected to the time when the last ball detected in the trajectory. In order to avoid to use such inaccurate data for calibration, the start time and end time of each trajectory are also stored and the trajectory reconstruction is done within this period of time. Note that the key idea is to use curve instead of line, which is used by Noguchi and Kato [4], for reconstructing trajectories. It is not the main purpose of this paper to find the best method for reconstructing the trajectory of the table tennis ball. Cubic spline is used as a simple method for trajectory fitting, and we compare the curve fitting to the linear fitting with respect to accuracy through the experiment.

4.3 Estimate temporal offset and fundamental matrix simultaneously Rough temporal offset τ_0 from a camera and the other camera, and ball trajectories in captured images were obtained until the previous step. We can estimate a corresponding ball position in a image with a ball position in the other image and rough temporal offset because trajectories are temporal function. A lot of corresponding points can be obtained and fundamental matrix can be computed from the corresponding points. However, the estimated fundamental matrix would not be accurate because of the error of τ_0. Let us denote fundamental matrix by $\boldsymbol{F}$, the ball position in the image captured with camera 1 at a time i by $\boldsymbol{p}_{1,i}$, the corresponding point of $\boldsymbol{p}_{1,i}$ in the image captured with camera 2 estimated from a temporal off-

(a) Camera 1 (b) Camera 2

Fig. 2. A frame captured in different two viewpoints. The background objects are eliminated because of the privacy reason.

set and a trajectory by $\boldsymbol{p}'_{2,i}$, the function that provides the distance by d, the function that provides median by Med. The error function $\mathrm{E}(\boldsymbol{F})$ is defined as following;

$$\mathrm{E}(\boldsymbol{F}) = \mathrm{Med}_i^N \, \mathrm{d}(\boldsymbol{p}_{1,i}, \boldsymbol{F}\boldsymbol{p}'_{2,i}) + \mathrm{d}(\boldsymbol{p}'_{2,i}, \boldsymbol{F}^T\boldsymbol{p}_{1,i}). \quad (2)$$

$\boldsymbol{F}\boldsymbol{p}'_{2,i}$ represents the epipolar line in the image captured with camera 1 that corresponding to $\boldsymbol{p}'_{2,i}$, $\boldsymbol{F}^T\boldsymbol{p}_{1,i}$ represents the epipolar line in the image captured with camera 2 that corresponding to $\boldsymbol{p}_{1,i}$. Median is used instead of average because median can be robust representatives even if false positives are included in the detection results. The estimation algorithm for temporal offset and fundamental matrix is described as following;

Step 1 Obtain corresponding points with temporal offset τ_j

Step 2 Compute fundamental matrix $\boldsymbol{F}_j$

Step 3 Seek optimal temporal offset τ_{j+1} that minimize the distance between every epipolar line, which is obtained with $\boldsymbol{F}_j$, and its corresponding point.

Step 4 Compute the error E_j and check if E_j is greater than the threshold. if true, add one to j and repeat from Step 1. if false, the loop is finished.

5. Experiment

5.1 Experimental condition A match played by two elite male players was taken with two cameras. A sample of the image taken with camera 1 is shown in Fig. 2(a) and a image taken with camera 2 is shown in Fig. 2(b). The kind of camera used in this experiment is Lumix GH3(Panasonic) and the resolution was set as 1920 × 1080, the frame rate was set as 60 fps. The 3D trajectory of the ten rallies, which were randomly chosen, were reconstructed. The temporal offset of the camera 2 was roughly estimated as 440 frames, it equals to 7333.333 ms, by observing with frame-by-frame playback and it was used as the initial temporal offset in this experiment.

First, a table tennis ball was detected in each frame by proposed method with the following parameters; U_r:13, L_r:4, L_i:30, L_c:0.5, U_v:9000, L_v:60. If the distance from manual recorded position is greater than 15 pixel, the detected positions was regarded as a false positive. If a ball can be recorded manually and no ball positions were successfully detected in a frame, the frame was regarded as a false negative. Next, trajectories in image space were reconstructed. The feature of this method is to use spline curve for estimating the motion of the ball between successive two frames. In order to evaluate the advantages of the feature, a linear method was conducted and the results was used as reference data. Let us name the proposed method by M_{spline}, the linear method by M_{line}. In this method, the each method was iterated 100 times and the result that provides the least error of fundamental matrix was selected. In addition to those, calibration was done by decomposing planar homography between table tennis court and image plane. Let us name the method by $\mathrm{M}_{homography}$. The results from the three methods were compared in the reprojection error.

Table 1. Detection results

	Camera 1	Camera 2
True positives	3014	2186
False positives	5	18
False negatives	542	1272
Recall	0.848	0.632
Precision	0.998	0.992

Table 2. Estimated temporal offset and reprojection error

	M_{spline}	M_{line}	M_{homo}
Temporal offset[ms]	7384.648	7384.772	-
Reprojection error[pixel]	1.08	11.15	46.59

5.2 Results The detection results were shown in Table. 1. The precision was high, greater than 0.99, and it is the biggest feature of this detection method. This result assures us that actual balls are the majority in the detected positions and many corresponding points were obtained correctly. The recall of camera 1 was greater than that of camera 2. Camera 1 is placed at higher position than camera 2(see Fig. 2) and that could be the main reason why such difference is occurred. This result implies that we can increase the number of detected ball positions by adjusting the camera positions.

The estimated temporal offset and the reprojection error are shown in Fig. 2. The methods that uses balls, M_{spline} and M_{line}, provides smaller reprojection error. 46.59 pixel is about 70 mm in this experiment. This value is close to the experimental result reported by Tamaki and Saito [7]. This experiment shows the method that calibrate cameras by decomposing planar homography would not provide accurate result compare to the methods that using balls.

M_{spline} provides smaller reprojection error than that of M_{line}. The difference between the two methods is only the method of interpolation, spline curve or line. A sample of the difference between spline interpolation and linear interpolation is shown in Fig. 3. The difference of them seems to be small. However, the error in the interpolation would make camera calibration inaccurate. Furthermore, both the error of the image position and the error of camera calibration would be merged in the triangulation. That's why small difference can strongly affect the reconstructed 3D coordinate. The reprojection error 1.08 pixel is originated from the error of de-

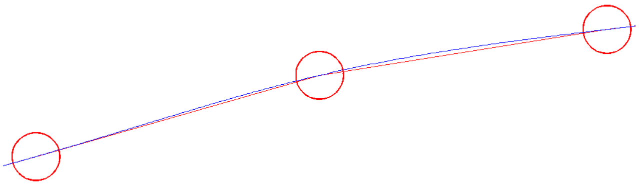
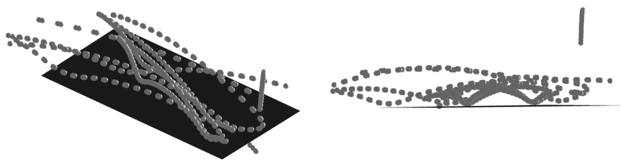


Fig. 3. Spline interpolation and linear interpolation. Red circles represents ball positions. Blue curves represents spline interpolation. Red lines represents linear interpolation.



(a) View 1

(b) View 2

Fig. 4. Reconstructed ball positions in 3-space. Black rectangle denotes a table tennis table. Gray spheres denote reconstructed ball positions

tected ball positions and the error of estimated temporal offset. This result indicates that the estimated temporal offset may be very close to accurate because of the very small reprojection error in M_{spline}, it is not evaluated quantitatively in this experiment though. From the results, proposing method is appropriate for the accurate measurement in a table tennis match. A sample of reconstructed ball positions in 3-space are shown in Fig. 4. It is experimentally demonstrated that the accurate measurement is able to be done by using proposing method in this paper.

6. Conclusion

A calibration method for performance analysis in table tennis is proposed in this paper. The method has the following features;

(1) It does not require time synchronized cameras.
(2) It requires only balls in a match for camera calibration.
(3) It detects a table tennis ball precisely, i.e. with few false positives.
(4) It calibrate cameras accurately, i.e. the reprojection error is about 1 pixel.

It is not difficult to automate the calibration if we use this method. This advantages might be used for improving the algorithm of ball detection because the geometrical relation between cameras provides epipolar lines and it helps to correspond points in two images.

References

(1) Y.Fei, K.Alexey, C.William, and K.Josef:"A Novel Data Association Algorithm for Object Tracking in Clutter with Application to Tennis Video Analysis", Proc. Computer Vision and Pattern Recognition, pp.634-641 (pub. 2006)
(2) Y.Huang, J.Llach, and C.Zhang:"A method of small object detection and tracking based on particle filters", Proc. International Conference on Pattern Recognition, pp1-4 (pub. 2008)
(3) I.L.Malagoli, R.D.Michele, and F.Merni:"A notational analysis of shot characteristics in top-level table tennis players", European Journal of Sport Science, ahead-of-print, pp.1-9 (pub. 2013)
(4) M.Noguchi and T.Kato:"Geometric and Timing Calibration for Unsynchronized Cameras Using Trajectories of a Moving Marker", IPSJ Transactions on Computer Vision and Image Media, Vol.47, No.5, pp.59-68 (pub. 2006)
(5) H.Numata, Y.Ujihara, H.Koyasu, H.Maekawa, T.Nagami, and K.Kanosue:"Estimation of Trajectory and Spin Analysis of Pitched Ball in Real Scene", Proc. FIT2009, Vol.8, No.3, pp.193-194 (pub. 2009)
(6) R.Hartley and A.Zisserman:"Multiple view geometry in computer vision", Cambridge Univ Press, pp.88 (pub. 2003)
(7) S.Tamaki and H.Saito:"Reconstruction of 3D Trajectories for Performance Analysis in Table Tennis", Proc. Computer Vision and Pattern Recognition Workshops, pp.1019-1026 (pub. 2013)
(8) K.Teachabarikiti, T.H.Chalidabhongse, and A.Thammano:"Players tracking and ball detection for an automatic tennis video annotation", Proc. of Control Automation Robotics & Vision, pp.2461-2494 (pub. 2010)
(9) X.Yu, C.H.Sim, J.R.Wang, and L.F.Cheong:"A trajectory-based ball detection and tracking algorithm in broadcast tennis video", Proc of International Conference on Image Processing, pp.1049-1052 (pub. 2004)
(10) C.Wei and Z.Yu-Jin:"Tracking Ball and Players with Applications to Highlight Ranking of Broadcasting Table Tennis Video", Proc. of Computational Engineering in Systems Applications, pp.1896-1903 (pub. 2006)
(11) Z.Zhang:"A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.11, pp.1330-1334 (pub. 2000)
(12) H.Zhang, W.Liu, J.J.Hu, and R.Z.Liu:"Evaluation of elite table tennis players' technique effectiveness", Journal of sports sciences, ahead-of-print, pp.1-9 (pub. 2013)
(13) Z.Zivkovic:"Improved adaptive Gaussian mixture model for background subtraction", Proc of International Conference Pattern Recognition, pp.28-31 (pub. 2004)

Sho Tamaki received the B.S., and M.S. degrees in physical education from Shizuoka University, Shizuoka, Japan, in 2007 and 2010, respectively. From 2010 to 2012, he worked at Japan Institute of Sports Sciences as a system engineer. Since 2012, he has been a Ph.D student of the Graduate School of Science and Technology, Keio University, Yokohama, Japan. He has been engaging in the research areas including sports science and computer vision.

Hideo Saito received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University,Yokohama, Japan, in 1987, 1989, and 1992, respectively. He has been on the faculty of Department of Electrical Engineering, Keio University, since 1992. From 1997 to 1999, he was a Visiting Researcher at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, where he joined the Virtualized Reality project. From 2000 to 2003, he was also a Researcher with PREST , JST. Since 2006, he has been a Professor in the Department of Information and Computer Science, Keio University. He has been engaging in the research areas including computer vision, image processing, augmented reality, and human-computer interaction. Dr. Saito is a member of IEICE, IPSJ, VRSJ, ITE, IIEEJ, and SICE, and a senior member of IEEE.