Three-Dimensional Baseball Strike Judgement by Monocular Video from Umpire's Viewpoint

Hiroto Fukuta
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
taketomohiro@akane.waseda.jp

Kosuke Nakayama
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
kosuke0013@fuji.waseda.jp

Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—In recent years, automatic judging is being introduced in many sports. In professional baseball, in particular, advanced equipment such as laser scanner and multiple cameras have been introduced. However, amateur and youth baseball teams cannot afford such expensive systems. Therefore, we propose a method to judge strike in three dimensions using a monocular video. The method consists of Camera Calibration block, Strike Zone Estimation block and Ball Trajectory Estimation block. Experiment results confirmed that automatic strike judgement in 3D space is possible by using the location information of the home base. Our method uses only one camera, reducing equipment cost and user effort.**

***Index Terms*—image processing, pose estimation, object detection, perspective projection transformation**

I. Introduction

In recent years, automatic judging is being introduced in many sports[1]. In baseball, the Automated Ball and Strike Calling System (ABS)[2] is being under consideration. This method can judge more accurate than humans by using specialized equipment. However, it requires multiple camera videos, and the equipment is expensive. Therefore, it's installation environment is limited.

In addition, the shortage of umpires has become an issue in youth and amateur baseball players. Accordingly, the umpire may be with no umpiring credentials or baseball experience. Unfortunately, it is impossible to use a method such as ABS in public ground where youth and amateur baseball games are played.

A method has also been proposed to obtain the ball's trajectory in three dimensions by using multiple camera videos[3]. However, this method requires at least three cameras for practical use. The use of multiple cameras leads to more complex processing and installation conditions.

Therefore, in this paper, we propose a method to judge strike in three dimensions from monocular images alone by using the location information of the home base. Our method is able to judge strikes by using only one monocular camera, so the cost of our method is also low. Thus, it has the potential to spread to many baseball players.

II. Related Work

A. OpenPose

OpenPose[4] is a method of pose estimation model that can estimate skeletal information of a person in video images. The detected skeletal information is output as the two-dimentional coordinates of the person's keypoints and the confidence level of the estimation. The number of keypoints detected depends on the dataset. In OpenPose, the XY coordinates originally start from the upper left corner of the screen. The confidence level ranges from 0 to 1, and the closer it is to 1, the more accurate the estimation is. It is set to 0 when keypoints fail to be detected. Also, OpenPose can process in real time.

B. Perspective Projection Transformation

The perspective projection transformation is a method of projecting an object from the real world onto a two-dimensional plane[5]. (X, Y, Z) is the three-dimensional coordinates of the real world coordinate system. (u, v) are the coordinates of the point projected onto the image plane. Also, f_x and f_y are the focal lengths of the camera, and the image center coordinates are (c_x, c_y). The internal parameter matrix A of the camera is given by

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

The external parameter matrix $[R|t]$ is calculated using the camera's rotation matrix and translation vector. It is given by

$$[R|t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}. \tag{2}$$

Moreover, the normalization parameter is s. The relationship between the three-dimensional coordinates in the real world, and the two-dimensional coordinates in the image plane is given by

979-8-3503-4017-4/23/$31.00 ©2023 IEEE

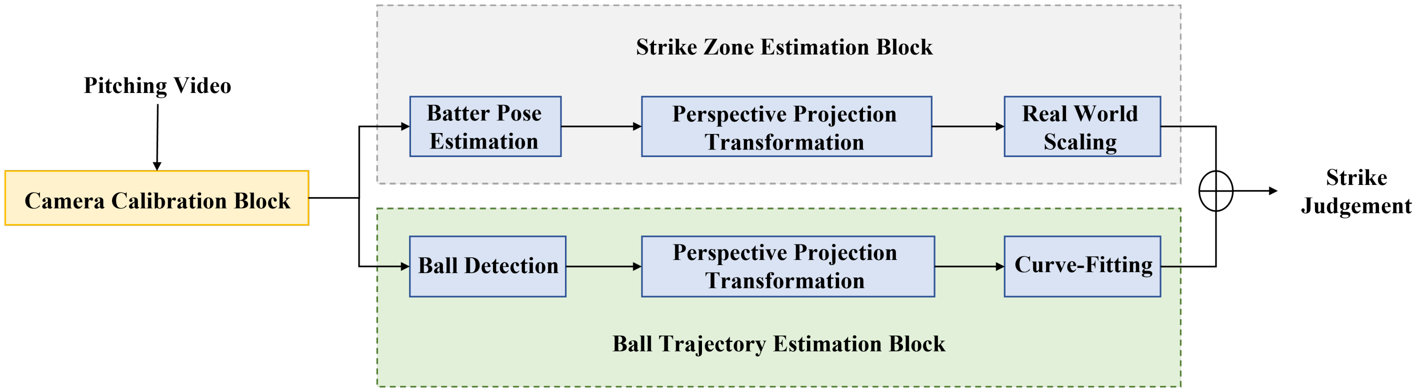

Fig. 1. Structure of the proposed method

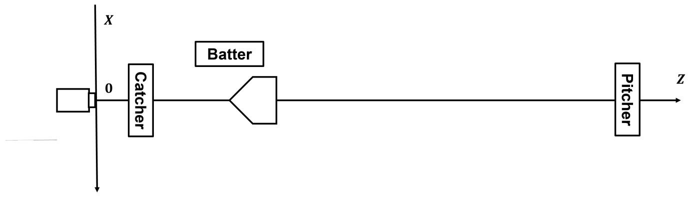

Fig. 2. Coordinate system from top view

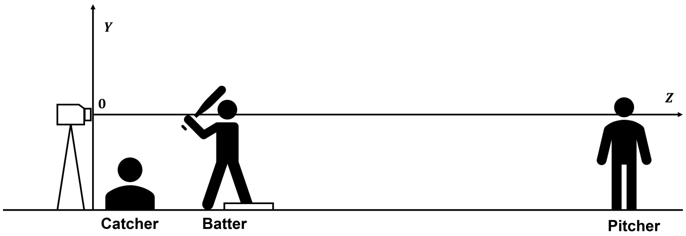

Fig. 3. Coordinate system from side view

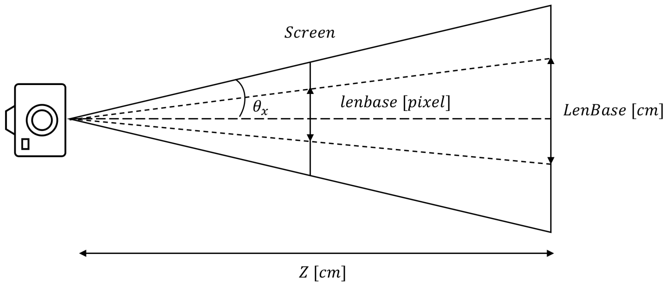

Fig. 4. Camera Calibration of X

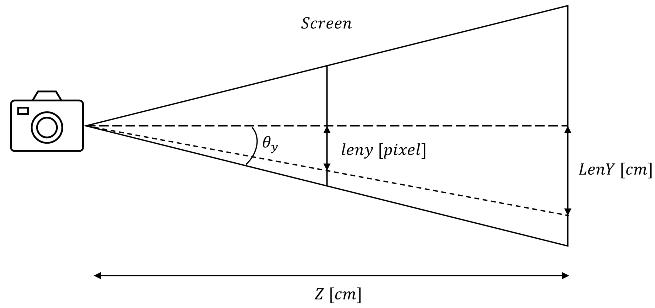

Fig. 5. Camera Calibration of Y

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (3)$$

III. Proposed Method

A. Structure of Proposed Method

The strike zone is a three-dimensional space above the home base, and varies for each batter[6]. The conventional method judges strikes in two dimensions using YOLOv5[7]: the strike zone and the ball trajectory. From the definition of strike zone, a two-dimensional strike judgment is not sufficient. Therefore, we propose a method to judge strikes in three dimensions from monocular videos by using the location information of the home base. On the other hand, our proposed method consists of three blocks: Camera Calibration block, Strike Zone Estimation block and Ball Trajectory Estimation block. Strike judgement is made by using the information obtained in the three blocks. The structure of the proposed method is shown in Fig.1.

B. Camera Calibration

In this paper, the camera's lens position is set as the origin. Fig.2. and Fig.3. show the coordinate system setup. Internal parameters of the camera are calculated using the location and size information of the home base. External parameters of the camera are calculated by measuring the camera position. Fig.4. and Fig.5. show the numerical value used for camera calibration. The camera's angles of view are θ_x and θ_y, the center coordinates of the screen are (c_x, c_y), and the distance from the camera to the home base is Z. The length of the home base in the screen is $lenbase$ and the actual length is $LenBase$. The height of the camera is $LenY$ and the length from the center of the screen to the home base is $leny$. The angle of view of the camera is given by

$$tan\theta_x = \frac{c_x LenBase}{Zlenbase}, \tag{4}$$

$$tan\theta_y = \frac{c_y LenY}{Zleny}. \tag{5}$$

Therefore, the internal parameter of the camera to be used are given by

$$A = \begin{bmatrix} \frac{c_x}{tan\theta_x} & 0 & c_x \\ 0 & \frac{c_y}{tan\theta_y} & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

In our method, the camera position is assumed to be fixed. For this reason, the external parameters of the camera can be calculated. The calculated internal and external parameters are used for the two blocks described below.

C. Strike Zone Estimation

First, pose estimation using OpenPose is performed for the batter. The top and bottom of the strike zone in the screen is given by

$$Top_{screen} = \frac{(RShoulder + LShoulder) + 2MidHip}{4}, \tag{7}$$

$$Bottom_{screen} = \frac{(RKnee + LKnee)}{2}. \tag{8}$$

The calculated two-dimensional coordinates of the strike zone are transformed into the three-dimensional coordinates by perspective projection transformation. In this case, the batter's standing position is assumed to be at the center of the batter's box. The distance from the camera to the batter is (cen_{bb}). Finally, the 3D coordinates of the strike zone are adjusted to a real-world scale. The real-world top and bottom used in the strike judgement are given by

$$Top_{real} = \frac{cen_{bb} tan\theta_y (Top_{screen} - c_y)}{c_y}, \tag{9}$$

$$Bottom_{real} = \frac{cen_{bb} tan\theta_y (Bottom_{screen} - c_y)}{c_y}. \tag{10}$$

D. Ball Trajectory Estimation

First, the ball coordinates are obtained from the pitching video using YOLOv5. Considering the effect of motion blur, the ball size in the screen is calculated based on the X coordinates. Then, the Z coordinates are calculated by comparing the ball size in the screen (BSS) with the real size of the ball ($RSB = 7.2cm$). Based on the calculated Z coordinates, the XY coordinates are transformed into three-dimensional coordinates by using perspective projection transformation. The XY coordinates of the ball's center in the screen are (cen_x, cen_y). The calculated XYZ coordinates are given by

$$Z = \frac{c_x RSB}{tan\theta_x BSS}, \tag{11}$$

$$X = \frac{Ztan\theta_x (cen_x - c_x)}{c_x}, \tag{12}$$

$$Y = \frac{Ztan\theta_y (cen_y - c_y)}{c_y}. \tag{13}$$

Finally, the three-dimensional coordinates of the ball are traced by curve fitting. Since the XY coordinates are dependent on the Z coordinates, curve fitting is done with a combination of XZ and YZ.

IV. Experiments

A. Whether or not video acquisition is possible

Our method uses pitching videos acquired from the umpire's viewpoint. There are two advantages to doing this. The first is that the camera taking the videos do not interfere with the game. The second is that the information necessary for judging strikes can be obtained more closely than before. However, the batter and pitcher events may interfere with the acquisition of the video from the umpire's viewpoint. Therefore, we check the effects on video acquisition of the main event patterns (Bunt, Habitual batting form, Underthrow, Sidethrow,Same-handed batter and pitcher and Opposite-handed batter and pitcher) performed by the batters and the pitchers.

Table 1 shows whether the video can be acquired or not from the umpire's viewpoint. It shows that our method can be confirmed that video acquisition is possible in all events. Since the umpires are standing in a position where they can judge strikes, it is considered that they are not affected by the events by the batter or the pitcher. Therefore, acquiring pitching videos from the umpire's viewpoint is effective.

B. Strike Judgement

Strike judgement is the binary classifications of ball and strike. We evaluate using the classification evaluation indices: Accuracy, Precision, Recall, and F-measure. The dataset used

TABLE I
WHETHER OR NOT VIDEO ACQUISITION FROM UMPIRE'S VIEWPOINT IS POSSIBLE.

Event	Can the video be acquired?
Bunt	Yes
Habitual batting form	Yes
Underthrow	Yes
Sidethrow	Yes
Same-handed batter and pitcher	Yes
Opposite-handed batter and pitcher	Yes

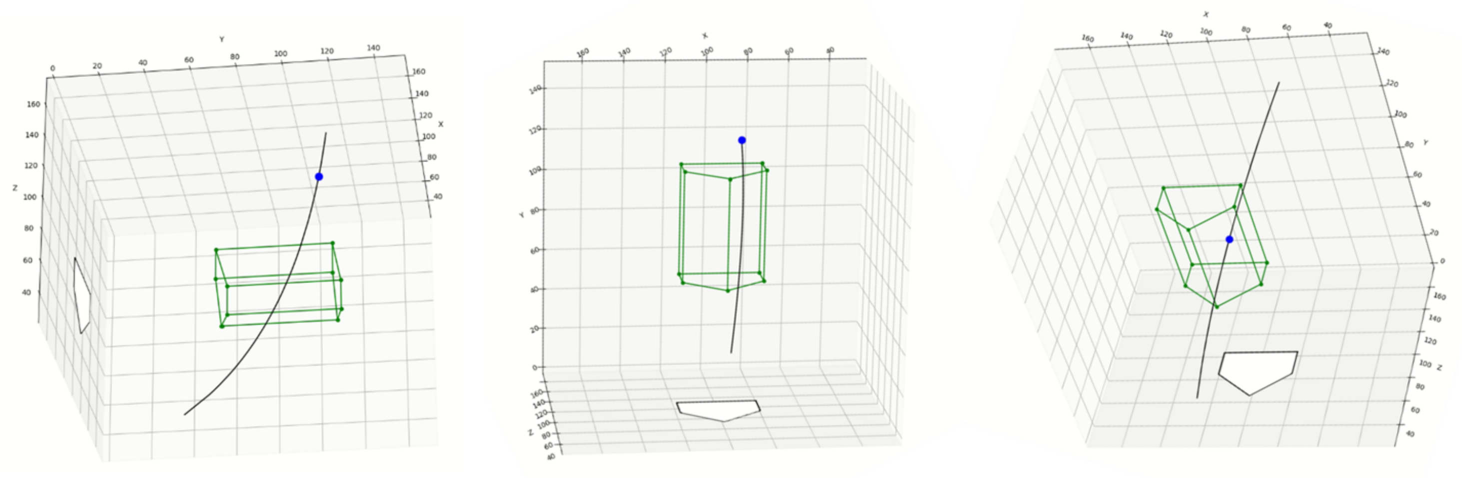

Fig. 6. Results in virtual 3D space

is 113 pitches obtained from the umpire's viewpoint. There are 53 pitches of strike judgements and 60 pitches of ball judgements in the dataset. Correct judgement was decided from the three viewpoints of the pitcher, batter, and catcher.

Table 2 shows the evaluation results of the three-dimensional strike judgements by using the proposed method. It shows that the proposed method is superior to the two-dimensional strike judgement method in all indices. The percentage of accuracy improved to nearly 90%, and recall improved by as much as 24%. However, the precision is as high as 98%, while the recall is as low as 77%. It shows that there is a tendency to judge strike-judged pitches as a ball. Therefore, it is necessary to ensure validity for balls that pass the edge of the strike zone.

Fig. 6. shows the composite results of the three-dimensional coordinate information obtained by using the proposed method. It confirmed that both ball trajectory and strike zone are reproduced in three dimensions. The black line represents the trajectory of the ball, and the blue sphere represents the position of the ball in a given frame.

Our method perceives the ball position and the strike zone to be changeable over time, allowing for the more accurate strike judgements. Also, it confirmed that the proposed method can obtain the three-dimensional information by the monocular video from the umpire's viewpoint.

TABLE II
CLASSIFICATION EVALUATION RESULT.

Evaluation Index	Conventional Method	Proposed Method
Accuracy	0.76	**0.88 (+0.12)**
Precision	0.93	**0.98 (+0.05)**
Recall	0.53	**0.77 (+0.24)**
F-measure	0.68	**0.86 (+0.18)**

V. CONCLUSION

In this paper, we proposed a new method to judge strikes. Specifically, we combined Strike Zone Estimation and Ball Trajectory Estimation by monocular video from umpire's viewpoint. From the experiments, we find that it is possible to get the pitching videos from the umpire's viewpoint. Moreover, the experiment result shows that it is superior to a typical two-dimentional strike judgements method in terms of strike judgment accuracy. We get the three-dimensional information by using the location information of the home base. Thus, our method is able to judge strikes by using only one monocular camera, so the cost of our method is also low. Therefore, it has the potential to spread to many baseball players.

REFERENCES

[1] B.T. Naik, M.F. Hashmi, and N.D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," Applied Sciences, vol.12, no.9, p.4429, Apr. 2022, doi: 10.3390/app12094429.

[2] Matthew Whitrock, "Developing MLB's Automated Ball/Strike System (ABS)," Published in MLB Technology Blog, 2021.

[3] S. Miyata, H.Saito, K.Takahashi, D. Mikami, M Isogawa, H. Kimata, "Ball-Tracking Based Temporal and Geometrical Camera Calibration and Reconstruction the 3D Trajectory for Baseball," IPSJ-CVIM, vol.2017-CVIM-205, no.40, pp.1-5, Jan. 2017.

[4] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Field, "IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp.1302-1310, Jul. 2017.

[5] Z. Zhang, "A flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.11, pp.1330-1334, Nov. 2000.

[6] Major League Baseball, "Strike Zone—Glossary—MLB.com," in MLB Official Website, https://www.mlb.com/glossary/rules/strike-zone 2023.

[7] Glenn Jocher, "YOLOv5 in pytorch," https://github.com/ultralytics/yolov5, Jun. 2020.