Ball 3D Trajectory Reconstruction without Preliminary Temporal and Geometrical Camera Calibration

Shogo Miyata Hideo Saito
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan
{miyata,saito}@hvrl.ics.keio.ac.jp

Kosuke Takahashi Dan Mikami Mariko Isogawa Hideaki Kimata
NTT Media Intelligence Laboratories
1-1 Hikarino-oka, Yokosuka, Kanagawa, Japan
{takahashi.kosuke, mikami.dan,isogawa.mariko,kimata.hideaki}@lab.ntt.co.jp

Abstract

This paper proposes a method for reconstructing 3D ball trajectories by using multiple temporally and geometrically uncalibrated cameras. To use cameras to measure the trajectory of a fast-moving object, such as a ball thrown by a pitcher, the cameras must be temporally synchronized and their position and orientation should be calibrated. In some cases, these conditions cannot be met, e.g., one cannot geometrically calibrate cameras when one cannot step into a baseball stadium. The basic idea of the proposed method is to use a ball captured by multiple cameras as a corresponding point. The method first detects a ball. Then, it estimates temporal difference between cameras. After that, the ball positions are used as corresponding points for geometrically calibrating the cameras. Experiments using actual pitching videos verify the effectiveness of our method.

1. Introduction

Computer vision-based sport assistance has been widely investigated using various methods [2], [4], [14], [16]. Our particular focus was on analyzing the trajectories of balls pitched in baseball. In baseball games, almost all plays start with the pitcher throwing the ball. The average number of balls the pitcher throws is 300 per game. Thus, analyzing ball trajectories can be a quite effective support method both for batters and pitchers.

In addition to analysis purposes, some commercial products tackle the problem of providing a virtual experience about the opposing pitcher before the game by using measured ball trajectories. To address this problem, EON Sports proposes a CAVE-based method and NTT proposes an HMD-based system. These are promising ways to provide sports training and can also be applied to entertainment purposes. Currently, commercially available sensors such as Trackman and PITCHf/x are used for professional baseball games. However, such systems are rather expensive and hard to carry to another cite. Ball trajectory measurement systems with portability significantly benefit scouting and recruiting. Thus, our target is a 3D ball trajectory measurement system with ordinary cameras that does not require calibration in advance.

The basic approach for measuring 3D trajectories from multiple cameras consists of two steps. First, it detects a ball at each camera image. Second, it measures the 3D position on the basis of triangulation. For this approach, some premises exist; shutter timing of the cameras is synchronized, and geometric relations such as relative position and orientation are known. However, these premises are difficult to be met when applying this approach to actual game situations. For example, to synchronize shutter of cameras, the signal generator and cameras should be connected, and thus camera setup is limited drastically. Another example is extrinsic calibration of cameras. To calibrate cameras extrinsically, an object with a known shape should be located on the view shared area. The main target in the work we describe in this paper is to overcome these difficulties.

To calibrate asynchronous cameras, Noguchi et al. used a marker. They simultaneously estimated differences in shutter timing and geometric information by using the trajectory of the marker. Tamaki et al. applied this method to table tennis. They used a ball during a rally substitute for the marker for asynchronous camera calibration.

Meanwhile, much computer vision research has been

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.26

conducted focusing on baseball [10],[11],[13]. In addition to the above mentioned commercially available 3D ball trajectory sensors, various methods such as Kalman filter-based object tracking have been applied to ball detection [3],[5],[12]. These studies mainly focus on ball analysis from one viewing angle, thus putting 3D ball trajectory reconstruction out of their focus. Gueziec et al. [7] used a camera deployed orthogonal to the batter's box to detect and reconstruct the 3D trajectory of the ball. However, this system relies heavily on camera setup, which can be done by using broadcasting systems.

This study tackles the problem of obtaining 3D ball trajectories under an easy setup, i.e., using an asynchronous and uncalibrated camera set. The main procedure of the proposed method is (1) ball detection in a 2D image sequence, (2) temporal and geometrical calibration on the basis of 2D ball trajectory, and (3) reconstructing 3D ball trajectory by triangulation.

The remainder of the paper is organized as follows: Section 2 introduces how the 3D ball trajectory is used in baseball. Sections 3 and 4 describe and evaluate the proposed method. We summarize key points in Section 5.

2. 3D Ball Trajectory in Baseball

As mentioned in Section 1, some commercial systems track ball 3D trajectory in baseball games. In this section, we first briefly introduce the system and then show some examples of how it is used.

One widely used ball tracking system is PITCHf/x, which is a camera-based ball tracking system. It is currently installed in all U.S. major league baseball stadiums. In addition to ball 3D trajectory, it provides various ball statistics, such as BRK, which represents how the ball bends, and PFX, which represents the spin derived deflection of the ball. Another ball tracking system is Trackman, which uses 3D Doppler radar. It provides a wide variety of information on topics such as the ball's spin rate, its spin axis, and how it bends.

Numerous ball trajectory analysis studies have been conducted on the basis of ball trajectory [1, 8, 15, 6]. Hamilton et al. applied machine learning technologies to PITCHf/x data and in doing so were able to predict the type of pitch that would be thrown before it was actually thrown [8]. Whiteside et al. analyzed performance changes in accordance with the number of innings that had been played on the basis of PITCHf/x data [15].

Recently, another way of using trajectory data has emerged, i.e., one that shows what the batter had experienced before. One example is a cave-based system provided by EON Sports. Another one is a HMD-based VR system that has been used by the Rakuten Eagles, a Japanese professional baseball team. They released the news they had obtained by installing a VR-based training system. These circumstances have increased the need to obtain 3D ball trajectory at the outside of the stadium.

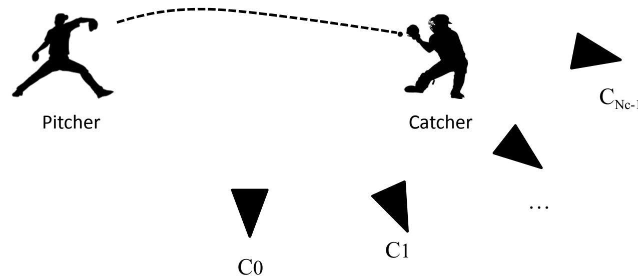

Figure 1. Multi-view capture system.

3. Proposed Method

In this section, we will describe the method we propose to estimate the 3D trajectory of a flying object such as a ball thrown by a pitcher from multi-view videos captured by unsynchronized and uncalibrated cameras.

Figure 1 illustrates the setup of our multi-view capture system. It consists of N_c static cameras $C_i (i = 0, \cdots, N_c - 1)$ that are not synchronized and not calibrated extrinsically. We assume that all cameras can observe the entire area of the field, that is, all the ball trajectories are in the field of view

We assume that all cameras can observe the all ball trajectories and their intrinsic parameters are estimated by using Zhang's method [17] beforehand.

Figure 2 shows the outline of the proposed algorithm. It first detects ball region candidates for all 2D images. Since these candidate regions include some misdetections, our proposed method removes them and estimates the 2D trajectory of the ball by fitting a uniformly accelerated motion model to the sequence of candidates of the ball region. Next, it calibrates the multiple cameras temporally and geometrically by using epipolar geometry. Finally, it estimates the 3D ball trajectory by using triangulation. The details of our algorithm are described below.

3.1. Detection of Candidates of Ball Region

In order to detect ball regions automatically, we introduce a frame subtraction approach. Let $f_i^j (i = 0, \cdots, N_c - 1, j = 0 \cdots, N_{f_i} - 1)$ denote the j th frame in an input video that has N_{fi} frames captured by C_i. We obtain a frame subtraction result $\bar{s}_i^j$ of the frame f_i^j as

$$\bar{s}_i^j = s_i^{j-1,j} * s_i^{j,j+1}, \tag{1}$$

where $s_i^{j-1,j}$ and $s_i^{j,j+1}$ are the frame subtraction results we respectively obtained by using the two frames (f_i^{j-1}, f_i^j) and (f_i^j, f_i^{j+1}).

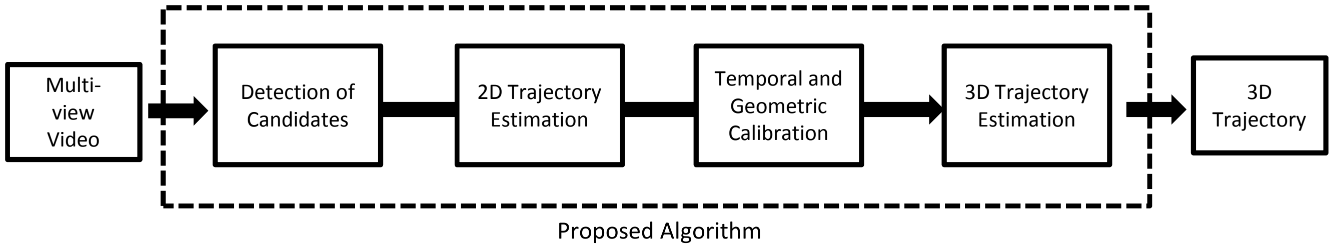

Figure 2. Outline of our algorithm.

The $\bar{s}_i^j$ value includes all change regions in time. We labeled these regions and added them to the ball region candidates $\mathrm{R}_i^j = \{r_{i:k}^j\}(k = 0, \cdots, |\mathrm{R}_i^j|-1)$ in the input video with f_i^j frames , where $|\mathrm{R}_i^j|$ denotes the number of candidates. Note that obvious non-ball regions such as those with too large an area, which represent the areas of a pitcher and a catcher, and too small an area, which represents observation noise, are removed beforehand by adapting a image filter which removes such area based on its size.

3.2. 2D Trajectory Estimation Based on Uniformly Accelerated Motion

In order to find the ball region from the candidates R_i^j for each frame, we focus on the sequential movements of the ball in time and take a RANSAC approach. We assume that the local movement of the ball can be modeled as uniformly accelerated motion in 2D images. In three sequential frames, we randomly chase one candidate from R_i^j per frame, such as $r_{i:k}^j, r_{i:k'}^{j+1}$ and $r_{i:k''}^{j+2}$, and compute the model parameters of uniformly accelerated motion from them. Let $\mathbf{a}_i^j$ and $\mathbf{v}_i^j$ denote the acceleration and velocity of the 2D ball trajectory in frame f_i^j and these parameters are computed as

$$\mathbf{a}_i^j = 2\frac{\delta t_j^{j+1}(\mathbf{c}_i^{j+2} - \mathbf{c}_i^{j+1}) - \delta t_{j+1}^{j+2}(\mathbf{c}_i^{j+2} - \mathbf{c}_i^{j})}{\delta t_j^{j+1}\delta t_{j+1}^{j+2}(\delta t_j^{j+1} + \delta t_{j+1}^{j+2})} \tag{2}$$

$$\mathbf{v}_i^j = \frac{\mathbf{c}_i^{j+1} - \mathbf{c}_i^j}{\delta t_j^{j+1}} - \frac{\mathbf{a}_j^i \delta t_j^{j+1}}{2} \tag{3}$$

$$\delta t_j^{j+1} = t_i^{j+1} - t_i^j \tag{4}$$

where t_i^j denotes the time of frame f_i^j, and $\mathbf{c}_i^j$ is the 2D coordinate of the centroid of $r_{i:k}^j$. Suppose that the time of f_i^j is t_0 and that the 2D position of the centroid of the ball on time t, *i.e.* $\hat{\mathbf{c}}_i^t$, is represented as

$$\hat{\mathbf{c}}_i^t = \hat{\mathbf{c}}_i^{t_0} + (t - t_0)\mathbf{v}_i^j + \frac{\mathbf{a}_i^j(t-t_0)^2}{2}. \tag{5}$$

Let $d(\mathbf{c}_{i:k}^t, \hat{\mathbf{c}}_i^t)$ denote the 2D distance between the centroid of one ball region candidate on time t and its estimated value be shown by Eq 5. If $d(\mathbf{c}_i^t, \hat{\mathbf{c}}_i^t)$ is smaller than threshold d_{th}, we regard $c_{i:k}^j$ as the *inlier* for the estimated parameters. By introducing this distance value, we define the evaluation function of the estimated parameters as

$$E = \Sigma_{j=i-N_W}^{i+N_W} \Sigma_{k=0}^{|\mathrm{R}_i^j|-1} \rho(\mathbf{c}_i^j) \tag{6}$$

$$\rho(\mathbf{c}_i^j) = \begin{cases} d^2(\mathbf{c}_i^j, \hat{\mathbf{c}}_i^j) & \text{if} \quad d(\mathbf{c}_i^j, \hat{\mathbf{c}}_i^j) < d_{th} \\ d_{th}^2 & \quad otherwise, \end{cases} \tag{7}$$

where N_W denotes the window size for the local area. In our algorithm, we selected one combination with the most inliers and the smallest value of 7 as the best combination of ball regions for modeling the 2D ball trajectory. Again, we computed the inliers with the best parameters and set them as the true 2D ball trajectories. Since these 2D ball trajectories are discrete values, we obtained continuous values for the entire 2D trajectories of the ball by applying a cubic spline interpolation to them.

3.3. Estimation of 3D Trajectory with Temporal and Geometrical Calibration

Let us assume that C_0 is the basis camera and that C_i has the time lag τ_i that satisfies $t_j^0 = t_j^i + \tau_i$. If τ_i is known, *i.e.* C_0 and C_i are synchronized, the fundamental matrix F_i between C_0 and C_i can be computed on the basis of epipolar geometry by utilizing the 2D trajectories of the ball in each camera as corresponding points. However, we assume that these cameras are casually set up and not synchronized. To address this problem, we focused on the epipolar geometry constraint and simultaneously estimated the ideal time lag τ_i and fundamental matrix F_i from the 2D ball trajectories

Let $\mathbf{p}_0^t$ and $\mathbf{p}_i^t$ denote the 2D positions on continuous 2D ball trajectories on time t in each camera. When synchronizing these cameras, $\mathbf{p}_0^t$ and $\mathbf{p}_i^t$ satisfy the epipolar constraint

$$\tilde{\mathbf{p}}_0^{t\top} F_0 \tilde{\mathbf{p}}_i^t = 0 \tag{8}$$

where $\tilde{\mathbf{x}}$ represents the homogeneous coordinate of $\mathbf{x}$. Using this constraint, we define the error function with a time lag τ_i as

$$E(\tau_i) = \frac{1}{N_t}\Sigma_{t=-N_t}^{N_t}\{D(\tilde{\mathbf{p}}_0^t, F_i(\tau_i)\tilde{\mathbf{p}}_i^{t+\tau_i})+ D(\tilde{\mathbf{p}}_i^{t+\tau_i}, F_i(\tau_i)^\top\tilde{\mathbf{p}}_0^t\} \tag{9}$$

where N_t denotes the evaluation range and $F_i(\tau_i)$ is the fundamental matrix estimated from the 2D trajectories with time lag τ_i. $D(\tilde{\mathbf{x}}, \mathbf{l})$ means the distance between a point $\mathbf{x}$ and a line $\mathbf{l}$. We obtain τ_i^* that minimizes Eq (9) as the appropriate time lag and obtain the fundamental matrix F_i^* with τ_i^*. Note that since corresponding two points from 2D ball trajectories sometimes do not satisfy the epipolar geometry due to some observation noise, we reject the outliers with LMedS when computing F_i^*.

The essential matrix E_i can be computed from the fundamental matrix F_i^* and the intrinsic camera parameters. This E_i can be decomposed to extrinsic parameters, such as rotation matrix and translation vector [9]. From these parameters, we estimate the 3D trajectory ball by DLT[9].

4. Experiment

To verify the effectiveness of the proposed method, we conducted 3D ball trajectory reconstruction from videos.

4.1. Experimental Setup

We captured baseball pitching with two cameras at an actual baseball stadium, reconstructing 3D ball trajectory reconstruction on the basis of the proposed method and that of the baseline method. For the baseline method, we employed the ball positions obtained by manual detection and the fundamental matrix calculated from 540 points of chessboard patterns placed between the pitcher's mound and home plate.

We used a Sony XDCam with resolution of 1280x720 pixels and 59.94 fps. Figure 3 shows the position of the two cameras and corresponding snapshots. Their shutter timings were synchronized. Intrinsic camera parameters for each camera were obtained by the method described by Zhang et al.[18].

We used ten pitches from the pitcher to the catcher to calibrate extrinsic camera parameters of cameras and reconstructed 3D ball trajectories. Though the shutter timings of the cameras were synchronized, we displaced two frames for one camera. In the following experiment, we examined time displacement at every 0.005 frame.

4.2. Ball Detection

Table 1 shows the detection results obtained for a moving object. In the table, "Only ball" means that the only moving object is the ball, "Ball+other object" means an object other than the ball was detected, and "Other object" means only objects other than the ball were detected. "Detection rate" is the ratio of the number of frames in which the ball region was detected among frames in which the ball exists. To determine whether the ball is detected or not, we used the position displacement between positions of the detected object and that of the manually detected ball; if the displacement was lower than 20 pixels, we judged that the ball was detected. As shown in Table 1, the proposed method failed to detect the ball in about 15% of the frames in each view.

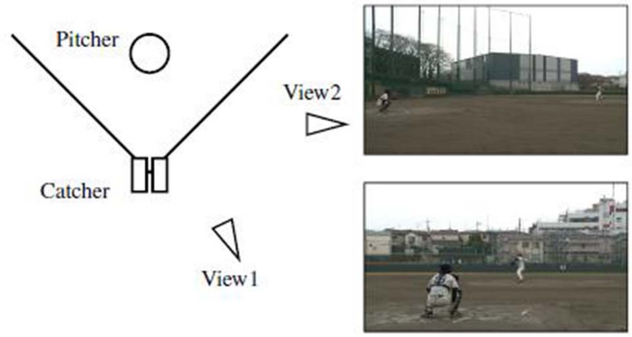

Figure 3. Position of cameras and corresponding images.

Table 1. Ratio of moving object detection.

	Cam1	Cam2
Only ball	61.6	84.8
Ball+other object	23.7	0.02
Other object	4.8	0.0
Detection rate	87.7	85.1

Table 2. Ball detection ratio.

	Cam1	Cam2
True positive	84.8	83.7
False positive	0.0	0.0
Detection rate	87.1	83.7

Table 2 shows the ball detection results. Using the ball trajectory restriction allowed us to correctly distinguish the ball from other moving objects. In this case the detection rate was slightly deteriorated by the ball trajectory fitting. We believe this is due to the trajectory model we used.

Moving object detection and ball detection are shown in Fig. 4. Black dots denote detected moving objects. Dots surrounded by cyan are detected as balls. Red circles denote manually provided ball positions.

4.3. Temporal and Geometrical Calibration

Estimated temporal displacement was 2.01 frames. In Fig. 5 we depict reprojection errors that occurred when displacement was changed, which shows that the proposed method conducts temporal calibration well. The evaluation result we got for the fundamental matrix is shown in Ta-

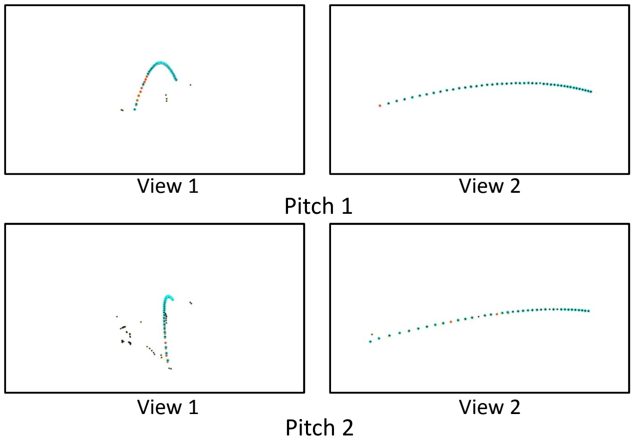

Figure 4. Detected moving object and detected ball.

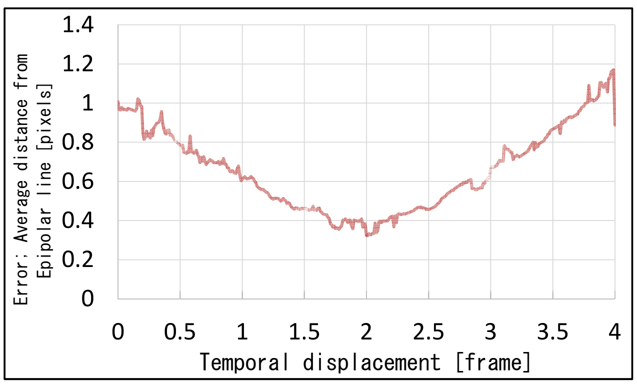

Figure 5. Temporal displacement estimation.

Table 3. Fundamental matrix evaluation.

	Baseline	Proposed
Distance from Epipolar line	0.962	1.13

ble 3; it shows the sum of reprojection errors, i.e., distances between corresponding points and epipolar line. For comparison purposes, we used the fundamental matrix obtained from manually provided ball positions. As shown in Table 3, the proposed method calculated the fundamental matrix with reprojection errors comparable to those obtained with the baseline method.

4.4. 3D Ball Trajectory Reconstruction

Figure 6 shows the results obtained for ball trajectory reconstruction. It shows positions of two cameras used for reconstruction and a reconstructed ball trajectory in 3D space. Camera position is shown by the combination of three lines that correspond to camera's light direction and horizontal axis and vertical axis. The blue lines show the results obtained with the proposed method and the red lines show those obtained with the baseline method. As the figure makes clear, the reconstructed trajectories of the two methods are duplicated, which means the proposed method well reconstructs 3D ball trajectories.

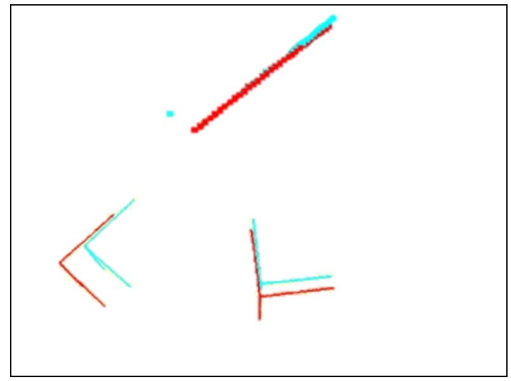

Figure 6. 3D reconstruction results.

5. Conclusion

This paper proposes a method for reconstructing 3D ball trajectories by using asynchronous and uncalibrated cameras. Our main idea is to use ball positions as corresponding points. We successfully reconstructed 3D ball trajectories by estimating them from videos and using the results obtained as corresponding points. In applying the proposed method to actually captured baseball pitching and comparing it with a baseline method, we verified that it accurately reconstructed the ball trajectories. Because our current experiments were limited, our future work includes verification in various setup.

References

[1] B. Baumer and D. Draghicescu. Mapping batter ability in baseball using spatial statistics techniques. In *Proc. Joint Statistical Meetings*, 2010.

[2] B. Chakraborty and S. Meher. A trajectory-based ball detection and tracking system with applications to shoot- ing angle and velocity estimation in basketball videos. In *Annual IEEE India Conference (INDICON)*, 2013.

[3] H.-T. Chen, H.-S. Chen, M.-H. Hsiao, W.-J. Tsai, and S.-Y. Lee. A trajectory-based ball tracking framework with visual enrichment for broadcast base- ball videos. *Journal of information Science and Engi- neering*, 2008.

[4] H.-T. Chen, W.-J. Tsai, S.-Y. Lee, and J.-Y. Yu. Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimedia Tools and Applications*, 2012.

[5] W.-T. Chu, C.-W. Wang, and J.-L. Wu. Extraction of baseball trajectory and physics-based validation for single-view baseball video sequences. In *ICME*, 2006.

[6] C. Dietrich, D. Koop, H. T. Vo, and C. T. Silva. Baseball4D: A tool for baseball game reconstruction & visualization. In *Proc. IEEE VAST*, 2014.

[7] A. Gueziec. Tracking pitches for broadcast television. *IEEE Computer*, 2002.

[8] M. Hamilton, P. Hoang, L. Layne, J. Murray, D. Padget, C. Stafford, and H. Tran. Applying machine learning techniques to baseball pitch prediction. In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*, 2014.

[9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[10] T. Hashimoto, Y. Uematsu, and H. Saito. Generation of see-through baseball movie from multi-camera views. In *IEEE International Workshop on Multimedia Signal Processing*, 2010.

[11] M. Takahashi, M. Fujii, and N. Yagi. Automatic pitch type recognition from baseball broadcast videos. In *IEEE International Symposium on Multimedia*, 2008.

[12] M. Takahashi, T. Misu, and M. Tadenuma. Real-time ball trajectory visualization using object extraction. *Visual Media Production*, 2005.

[13] C. Theobalt, I. Albrecht, J. Haber, M. Magnor, and H.-P. Seidel. Pitching a baseball: tracking high-speed motion with multi-exposure images. *ACM Transactions on Graphics*, 2004.

[14] X.-F. Tong, H.-Q. Lu, and Q.-S. Liu. An effective and fast soccer ball detection and tracking method. *Proceedings of the 17th International Conference on Pattern Recognition*, 2004.

[15] D. Whiteside, D. N. Martini, R. F. Zernicke, and G. C. Goulet. Changes in a starting pitcher's performance characteristics across the duration of a major league baseball game. *International Journal of Sports Physiology and Performance*, 2016.

[16] X. Yu, Q. Tian, and K. W. Wan. A novel ball detection framework for real soccer video. In *IEEE International Conference on Multimedia and Expo*, 2003.

[17] Z. Zhang. A flexible new technique for camera calibration. *TPAMI*, pages 1330–1334, 2000.

[18] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 2000.