Event-based Ball Spin Estimation in Sports

Takuya Nakabayashi
Keio University
Yokohama, Japan
nakka0204@keio.jp

Kyota Higa
NEC Corporation
Kawasaki, Japan
k-higa@nec.com

Masahiro Yamaguchi
NEC Corporation
Kawasaki, Japan
masahiro-yamaguchi@nec.com

Ryo Fujiwara
NEC Corporation
Kawasaki, Japan
r1-fujiwara@nec.com

Hideo Saito
Keio University
Yokohama, Japan
hs@keio.jp

Abstract

Ball spin estimation in sports is important for analyzing the game. Since spin is generally too fast to be captured by a conventional camera, a high-speed camera is often used to capture images of the ball and estimate its spin. However, since a high-speed camera is not robust to changes in the lighting conditions, it is difficult to estimate spin in some environments. To solve these problems, this paper proposes a new method for ball spin estimation using an event camera. An event camera is a sensor inspired by the visual system of animals, which outputs the brightness changes in a scene. Event cameras have advantages such as high temporal resolution and high dynamic range, and can accurately capture the motion of a fast-spinning ball in various lighting conditions. Experimental results in a synthesized dataset showed that the proposed method can stably estimate spin up to 500 rps. It is also confirmed that the proposed method can estimate spin in the data obtained from actual sports games.

1. Introduction

The analysis of ball motion in ball games has been one of the most popular research topics in sports analysis. Estimating the trajectory, velocity, and spin provides valuable insights for analyzing player performance and game strategy.

In this paper, we focus on spin estimation in ball motion analysis. Spin plays a crucial role in ball games. In baseball, for example, spin alters the trajectory of the ball, making it more challenging for the hitter to make contact. Similarly, in table tennis, the optimal racket angle for receiving depends on spin, necessitating players to accurately estimate spin of the opponent's shots. Thus, spin makes a game more strategic.

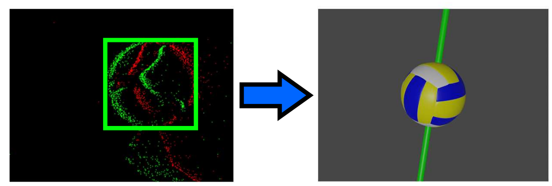

Figure 1. Spin estimation from events.

Ball spin estimation is essential for analyzing players' performance. Spinning a ball is a fundamental skill in various ball games, and quantitatively measuring the spin axis and angular velocity is crucial for evaluation. For instance, Trackman is among the instruments used to measure spin in golf and baseball, widely used among professional players for performance analysis.

Many methods for estimating spin from video have been proposed. Spin in sports can exceed 100 rps, making it difficult to accurately capture with ordinary cameras. Therefore, in the existing method, the ball is captured with a high-speed camera, and the position and movement of markers and logos on the ball are used as cues to estimate its pose and spin. However, existing methods using high-speed cameras have limitations. The exposure time of a high-speed camera is shorter than that of an ordinary camera, requiring a sufficiently bright lighting conditions. This presents a practical issue as inference performance may be significantly affected by the lighting conditions.

We propose a new method for ball spin estimation using an event camera. The event camera is a sensor inspired by the visual system of animals, which outputs information on pixel luminance changes as events. Compared to conventional cameras, the event camera offers several advantages including high temporal resolution, high dynamic range, high data efficiency, and low power consumption. In particular, spin of a ball in sports is very fast, making it highly effective to utilize an event camera with high temporal resolution. Additionally, the high dynamic range of the event camera allows for the estimation independent of the lighting environment.

We estimate spin using Contrast Maximization framework [4], which estimates motion parameters from events. Since this method does not depend on the type of ball, it can be applied to basically any sport. In addition, by using events characterized by high dynamic range and high temporal resolution as input, it is possible to estimate the spin of a high-speed ball under a variety of lighting conditions.

We conducted experiments to verify the effectiveness of the proposed method in spin estimation and confirmed that it outperforms the image-based method quantitatively. Experiments on a synthetic dataset demonstrate that the proposed method can consistently estimate spin up to 500 revolutions per second (rps). Additionally, experiments conducted using data collected in laboratory settings and during sports games confirm that the proposed method can accurately estimate spin in real-world scenarios.

The contributions of this paper are as follows:

- We propose a new method for estimating spin using an event camera. By using an event camera, it is possible to accurately measure spin of a ball at extremely high speeds.
- The proposed method is general-purpose because it can estimate spin of balls used in various sports. Experimental results show that the proposed method successfully estimates spin of volleyballs and soccer balls.
- We conducted experiments on synthetic datasets and quantitatively confirmed that the proposed method can estimate spin more accurately than image-based methods. The experiments show that the proposed method can stably estimate spin up to 500 rps.
- We conducted experiments on data captured in the laboratory and in actual volleyball matches, and showed that the proposed method is applicable to real-world data.

2. Related Work

2.1. Ball Spin Estimation

Spin estimation can be classified into two categories. The first is the direct approach, which estimates spin by directly observing the logo or texture on the ball and the trajectory of the ball. The second is the indirect approach, in which the player's pose and the racket movement are used as cues for estimation.

Direct approach There are two types of the direct approach: texture-based and trajectory-based.

Texture-based methods estimate the pose and angular velocity of a ball by analyzing the position and movement of a texture on the ball's surface. This texture could be a preexisting logo or pattern on the ball, or a specially applied pattern designed for measurement purposes. Zhang et al. [17] proposed a method to estimate the pose by tracking the position of a logo mark on a table tennis ball from the difference images. Tebbe et al. [16] also employ a similar approach to estimate angular velocity by detecting the position of a logo mark. While these methods excel in using the official ball itself and can thus be utilized in official games, they encounter challenges in estimating the angular velocity when the camera loses sight of the logo due to changes in the ball's orientation. Furuno et al. [3] introduced a technique involving colored lines drawn on the ball's equator and perpendicular great circle, estimating spin from the intersections of these lines. Similarly, Gossard et al. [6] estimated the ball's orientation by applying a specific dot pattern to the ball surface, tracking its position using Convolutional Neural Network (CNN), and employing the Kabsch algorithm. While these methods can estimate the pose of the ball as long as the markers are visible to the camera, they cannot be applied directly to the official ball because of the need to add patterns to the ball, making it impossible to measure the pose in official games.

The trajectory-based method estimates spin of a ball by analyzing changes in its trajectory resulting from its spin. When a spinning ball moves through the air, it experiences a force determined by the direction and velocity of its spin, known as the Magnus effect, which alters its trajectory. This phenomenon is akin to the curve observed in a baseball pitch. The trajectory-based method tracks the trajectory of a ball by continuously estimating its position in the image and then calculates spin based on physical principles. Chen et al. [2] and Su et al. [14] utilize stereo vision to acquire the 3D trajectory of the ball, incorporating spin into their predictions using a physical model based on fluid dynamics.

Balls in ball games often move at high speeds, which causes motion blur when captured by ordinary cameras. Therefore, many methods use a high-speed camera to capture the ball in order to minimize the effect of motion blur. Generally, a high-speed camera is susceptible to the lighting environment in that the exposure time is short and sufficient illumination is required during the shooting. In addition, if the frame rate of the camera is insufficient, the motion of the ball between adjacent frames may be too large to estimate the fast spin.

To address the limitations of the high-speed camera method, we propose a novel approach for estimating the rotation of a ball utilizing an event camera. Event cameras offer a high temporal resolution on the order of microseconds and are immune to motion blur. Moreover, their high

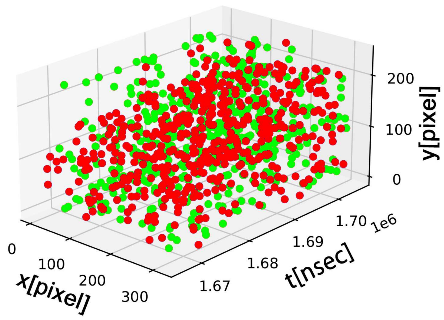

Figure 2. Example of events. The color of the dots corresponds to the polarity of the event. Events are spatio-temporally sparse point cloud-like data.

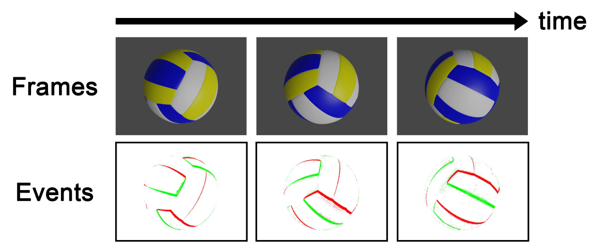

Figure 3. Events that occur when the ball spins. Events are concentrated around the edges on the image.

dynamic range enables operation in diverse lighting conditions.

Indirect approach The indirect approach utilizes the player's posture and racket movement as cues to estimate spin. Sato and Aono [12] employed a DNN model to classify strokes based on estimated joint positions. Blank et al. [1] classified players' strokes using data gathered by an IMU attached to a table tennis racket. While these methods can roughly classify strokes, accurately estimating details such as the spin axis and angular velocity of the ball proves challenging.

2.2. Event-based Motion Estimation

Due to its high temporal resolution, the event camera can accurately capture high-speed motion that would cause motion blur in conventional cameras. Unlike images, events are sparse data and conventional image-based algorithms cannot be applied directly to them. Several algorithms [4, 7, 10, 13] have been proposed for estimating information about motion in a scene (translational velocity, angular velocity, optical flow, etc.) from events. The proposed method extends the Contrast Maximization framework to spin estimation.

3. Preliminaries

3.1. Event Camera

An event camera is a sensor that outputs luminance changes as events. Unlike conventional cameras, which synchronously acquire luminance at all pixel positions at fixed time intervals, event cameras output events asynchronously at each pixel position only when the luminance changes. Therefore, the events are spatio-temporally sparse data, as illustrated in Figure 2. An event e can be expressed as $e = (\boldsymbol{x}, t, p)$, where $\boldsymbol{x} = [u, v]$ represents the pixel coordinates, t denotes time, and p indicates polarity. p can be either +1 or -1 depending on whether the luminance is greater or less.

As events are triggered by luminance changes in the scene, they predominantly occur along edges, where spatial luminance changes significantly. Figure 3 illustrates the distribution of events over a certain period of time, demonstrating their concentration around moving edges. In essence, events encode the motion of objects within the scene.

3.2. Contrast Maximization Framework

We utilize Contrast Maximization framework [4] to estimate the spin axis and angular velocity from events. Contrast Maximization framework is a method used to estimate parameters associated with scene motion from events.

The Contrast Maximization framework estimates the motion parameter θ from N_e events $\{e_k\}_{k=1}^{N_e}$. Each event is represented as $e_k = (\boldsymbol{x}_k, t_k, p_k)$. Let $\boldsymbol{x}_k$ be warped to the reference time t_{ref} by θ. $\boldsymbol{W}$ is the motion model, and $\boldsymbol{x}'_k$ be the coordinates after warping, expressed as follows:

$$\boldsymbol{x}'_k = \boldsymbol{W}(\boldsymbol{x}_k, t_k; \theta). \tag{1}$$

For example, if a constant velocity linear motion is assumed for W, then

$$\boldsymbol{W}(\boldsymbol{x}_k, t_k; \theta) = \boldsymbol{x}_k + (t_{ref} - t_k)\theta. \tag{2}$$

In this case, θ denotes the velocity. The motion model $\boldsymbol{W}$ warps all events and creates an Image of Warped Events (IWE) by adding the polarity p_k to the warped coordinates $\boldsymbol{x}'_k$. The value of IWE at pixel position $\boldsymbol{x}$ is calculated as follows:

$$H(\boldsymbol{x}; \theta) = \sum_{k=1}^{N_e} p_k \delta(\boldsymbol{x} - \boldsymbol{x}'_k), \tag{3}$$

where δ is defined as follows:

$$\delta(\boldsymbol{x}) = \begin{cases} 1 & |\boldsymbol{x}| = 0 \\ 0 & otherwise. \end{cases} \tag{4}$$

As mentioned above, since events follow points on the edge, all events generated from the same point will be warped to the same position if the events are warped with the correct motion paramete. In this case, the variance of IWE is maximized. Assuming that the value of IWE at coordinate (i, j)

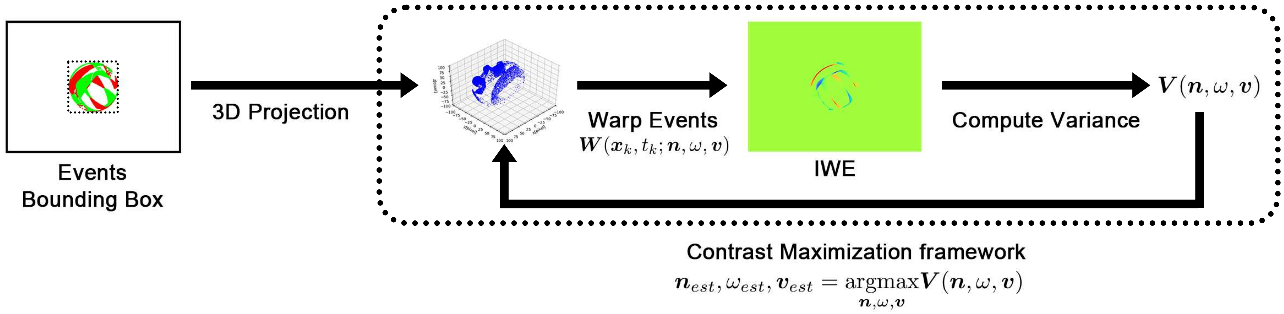

Figure 4. Overview of the proposed method. The proposed method uses the event and ball bounding box as input and normalized spin axis vector, angular velocity, and translational velocity as output. The proposed method first projects the events into a three-dimensional space by calculating the depth. Next, the parameters are estimated by iterative optimization using the Contrast Maximization framework.

is h_{ij} and the mean value of IWE is μ_H, the variance $V(\theta)$ is calculated as follows:

$$V(\theta) = \frac{1}{N_p} \sum_{i,j} (h_{ij} - \mu_H)^2, \tag{5}$$

where N_p denotes the total number of pixels in IWE. The Contrast Maximization framework estimates motion by optimizing the parameter θ to maximize $V(\theta)$. Therefore, the Contrast Maximization framework is expressed as follows:

$$\theta_{est} = \underset{\theta}{\operatorname{argmax}} V(\theta). \tag{6}$$

4. Method

4.1. Problem Definition

Our method takes as input the bounding box of the ball at time t_0 and an event sequence $\{e_k\}_{k=1}^{N_e}$ that have occurred since t_0, and outputs the normalized spin axis vector $\boldsymbol{n} = [n_x, n_y, n_z]^T$ in the camera coordinate system, the angular velocity ω, and the translational velocity $\boldsymbol{v} = [v_x, v_y]^T$.

Our method assumes that the size of the ball on the image remains constant during the estimation process. Strictly speaking, the size of the ball changes due to its translational motion, but this change is considered negligible since the input event sequence is very short in time.

4.2. 3D Projection

Since the input events include events that occur on backgrounds other than the ball, we first extract only events that occur on the surface of the ball. Given the bounding box, we extract only the events that occur inside the bounding box.

Although $\boldsymbol{x}_k$ represents two-dimensional coordinates, we initially compute the depth corresponding to each event to estimate the three-dimensional spin of the ball. The method for determining depth is based on the approach proposed by Tamaki et al. [15]. The position $\boldsymbol{c_t} = [u_t, v_t]$ of

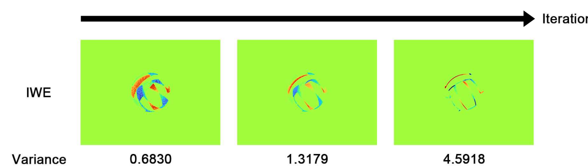

Figure 5. The process of IWE optimization. As the optimization progresses, the warp location of the event is concentrated on the edge of the ball, and the value of variance increases.

the ball center on the image at time t is expressed as follows:

$$\boldsymbol{c}_t = \boldsymbol{c}_{t_0} + \boldsymbol{v}(t - t_0). \tag{7}$$

Assuming that the ball is a perfect sphere of radius r, for a point (x, y, z) on the ball surface at time t, we have

$$(x - u_t)^2 + (y - v_t)^2 + z^2 = r^2. \tag{8}$$

Thus, the depth z_k corresponding to the event e_k is expressed as follows:

$$z_k = \pm\sqrt{r^2 - (x_k - u_{t_k})^2 - (y_k - v_{t_k})^2}. \tag{9}$$

If the camera is oriented toward the positive direction of the z-axis, it observes the $z < 0$ side of the sphere. Therefore, we utilize the negative value of Equation 9 as the depth.

4.3. Event Warping

The operation of spinning a point around an arbitrary spin axis can be concisely expressed using quaternions. Consider an event e_k projected in 3-dimensional space around a normalized spin axis vector $\boldsymbol{n} = [n_x, n_y, n_z]^T$ and warped to a reference time t_{ref} with the angular velocity ω. Let $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ denote the imaginary units in the quaternion. Since the magnitude of the spin angle corresponding to the event e_k is $\phi = \omega(t_{ref} - t_k)$, the quaternion $\boldsymbol{q}_k$ corresponding to

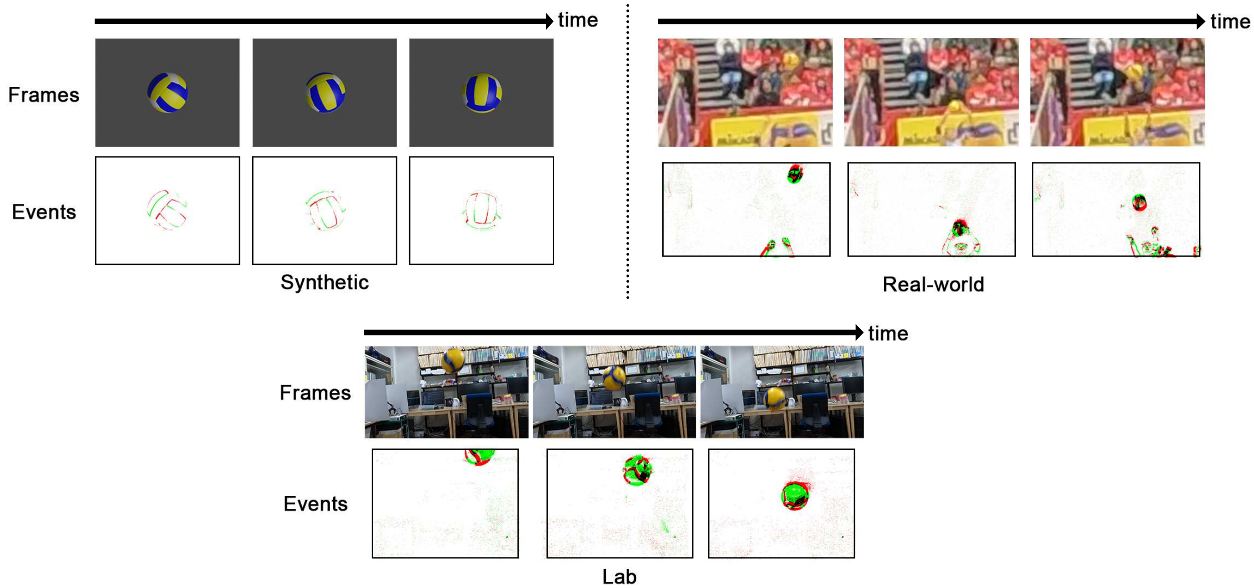

Figure 6. Comparison of events with frames in Synthetic Dataset, Lab Dataset, and Real-world dataset. Events that occurred within a certain period of time were accumulated and visualized. Green dots correspond to events with $p > 0$ and red dots to events with $p < 0$.

spin is expressed as follows:

$$\boldsymbol{q}_k = n_x \sin\frac{\phi}{2}\boldsymbol{i} + n_y \sin\frac{\phi}{2}\boldsymbol{j} + n_z \sin\frac{\phi}{2}\boldsymbol{k} + \cos\frac{\phi}{2}. \quad (10)$$

Let $\boldsymbol{p}_k$ and $\boldsymbol{p}'_k$ be the quaternions corresponding to the events before and after the spin, respectively. The relationship between $\boldsymbol{p}_k$ and $\boldsymbol{p}'_k$ is expressed as follows:

$$\boldsymbol{p}'_k = \boldsymbol{q}_k \boldsymbol{p}_k \bar{\boldsymbol{q}}_k, \quad (11)$$

where $\bar{\boldsymbol{q}}_k$ denotes the quaternion conjugate of $\boldsymbol{q}_k$.

4.4. Optimization

As discussed in Section 3.2, the Contrast Maximization framework optimizes the motion parameters to maximize the variance of the IWE, defined by Equation 5. To compute the IWE, we require the warp position of an event in two dimensions on the image plane. To project the 3D warp positions of the events computed in Sections 4.2 and 4.3 onto the image plane, we simply exclude the event depth. For a more precise computation of the projection position on the image plane, it is necessary to utilize the pinhole camera model, which requires camera calibration. However, excluding the event depth from the projection computation does not necessitate knowledge of the camera's internal parameters. Therefore, we adopt this simplified method.

Any method (e.g. steepest descent, Adam [8]) can be used for parameter optimization. Figure 5 illustrates that as the optimization progresses, the projection position of events in IWE gradually concentrates around the edges of the ball, leading to an increase in variance.

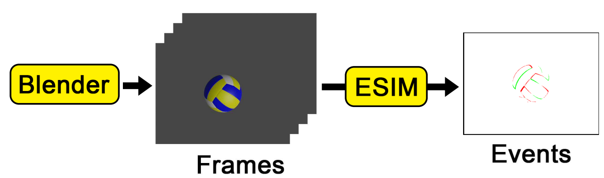

Figure 7. Dataset synthesis. The frame rendered in Blender is input to ESIM to generate events.

5. Experiment

In this paper, we conducted experiments using three datasets, Synthetic Dataset, Lab Dataset, and Real-world Dataset, to verify the effectiveness of the proposed method.

5.1. Implementation Details

In our experiments, we utilized $N_e = 50000$ events for a single spin estimation, with the Contrast Maximization framework running for 1500 iterations. For optimization, we employed Adam.

5.2. Experiment on Synthetic Dataset

Synthetic Dataset. Synthetic Dataset comprises frames and events synthesized through simulation, facilitating the quantitative evaluation of method performance as the ground truth, including spin axis vectors and angular velocities, is known.

The dataset features a series of balls rotating around specific axis without translational motion, encompassing two

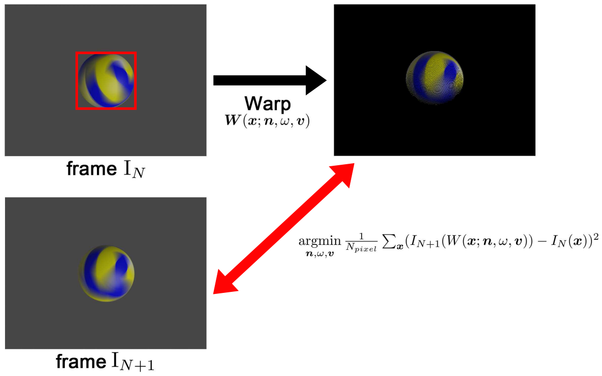

Figure 8. Overview of the image-based baseline method. It warps the pixels on the ball and optimize to minimize the MSE between the warped pixels and pixels on the adjacent frame.

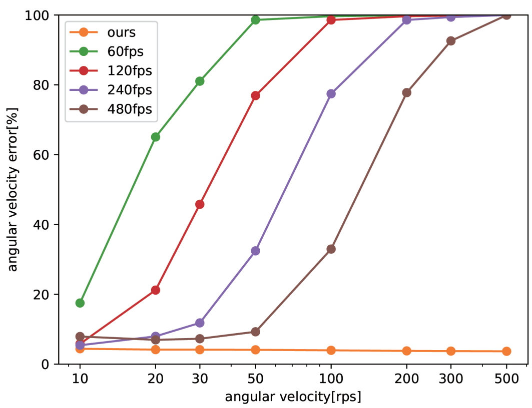

Figure 9. Angular velocity errors on synthetic volleyball dataset. Unlike image-based baselines, our method shows little change in error rate as angular velocity increases.

types: volleyball and soccer ball. The volleyball dataset encompasses data for eight angular velocities: 10, 20, 30, 50, 100, 200, 300, and 500 rps, each comprising 500 sequences with randomly determined spin axis vectors. The soccer ball dataset comprises 500 sequences with an angular velocity of 100 rps. Both events and images have a resolution of 640 pixels $\times$ 480 pixels.

The dataset synthesis procedure is illustrated in Figure 7, employing Blender and ESIM [11]. Blender is a software designed for 3D computer graphics, utilized to render the spinning ball. ESIM, an event camera simulator, was employed to generate events from the sequence of frames synthesized in Blender. ESIM is adept at synthesizing events from video sequences. The threshold of event occurrence in ESIM (contrast_threshold_pos and contrast_threshold_neg) is set to 0.15.

Baselines. In this experiment, we implemented an image-

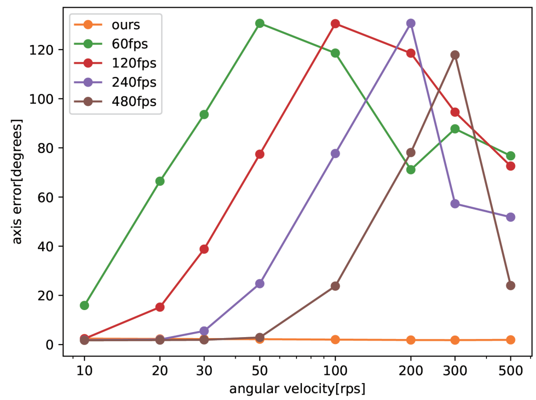

Figure 10. Axis errors on synthetic volleyball dataset. Our method can accurately estimate the spin axis independent of angular velocity.

based baseline method to evaluate the effectiveness of using an event camera in spin estimation and compared its performance. Figure 8 provides an overview of the baseline method, which is based on the approach proposed by Tamaki et al. [15]. This method takes two adjacent frames, I_N and I_{N+1}, from the video as input and estimates the spin axis vector $\boldsymbol{n}$, angular velocity ω, and translational velocity $\boldsymbol{v}$. The baseline method warps the ball surface pixels in I_N with the estimated parameters and optimizes these parameters to minimize the Mean Squared Loss when compared with the pixel values in I_{N+1}. The input to the baseline method comprised images generated at four different shutter speeds: 60, 120, 240, and 480 fps. For optimization, we utilized Adam with 1500 iterations.

Evaluation metrics. Two evaluation metrics, axis error and angular velocity error, were utilized. The axis error measures the error in the orientation of the spin axis vector and is calculated as the angle between the estimated spin axis vector and the spin axis vector of the ground truth. The angular velocity error quantifies the error rate of the estimated angular velocity.

Results. Figures 9 and 10 evaluate the performance of both the proposed method and the baseline method using volleyball data from Synthetic Dataset. The proposed method consistently exhibits small errors regardless of the magnitude of the angular velocity. Conversely, the error of the baseline method increases rapidly as the angular velocity of the input data rises. This is because the motion of the ball between adjacent frames increases as the angular velocity increases, making optimization difficult, and the motion blur in the frames is so severe that the texture on the ball becomes blurred. Given the event camera's exceptionally high temporal resolution, motion blur is effectively eliminated, circumventing this issue.

Figure 11. Device setup. On the left is the GoPro HERO 10 and on the right is Prophesee's Evaluation Kit.

Table 1 shows the results of evaluating the performance of the proposed method on 500 sequences of two types of balls (volleyball and soccer ball) spinning at 100 rps. Remarkably, the proposed method accurately estimates ball spin even when the ball type changes. Unlike many existing methods, the proposed approach can estimate spin independently of the ball's texture, making it applicable across various sports.

Table 1. Quantitative results for different types of balls.

	volleyball	soccer
axis error [degrees]	2.00	1.47
angular velocity error [%]	3.95	4.12

5.3. Experiment on Lab Dataset

Lab Dataset. Lab Dataset comprises sequences captured in the laboratory using real event cameras. It was assembled to qualitatively evaluate the proposed method's performance in a relatively noisy environment. Both volleyball and basketball sequences were recorded as they rotated and fell using the Prophesee Evaluation Kit 3 and GoPro HERO 10 in a stereo configuration, as illustrated in Figure 11. The initial position of the ball is provided through annotation.

Results. Figure 12 illustrates the visualization results of the estimated spins on Lab Dataset. Upon comparing the textured motion of the ball in the video frames with the estimated spin visualization results for both volleyball and basketball, the motions closely align, suggesting that the proposed method offers a reasonable estimation. Consequently, it appears that the proposed method is capable of accurately estimating the spin of various types of ball sequences captured by real event cameras.

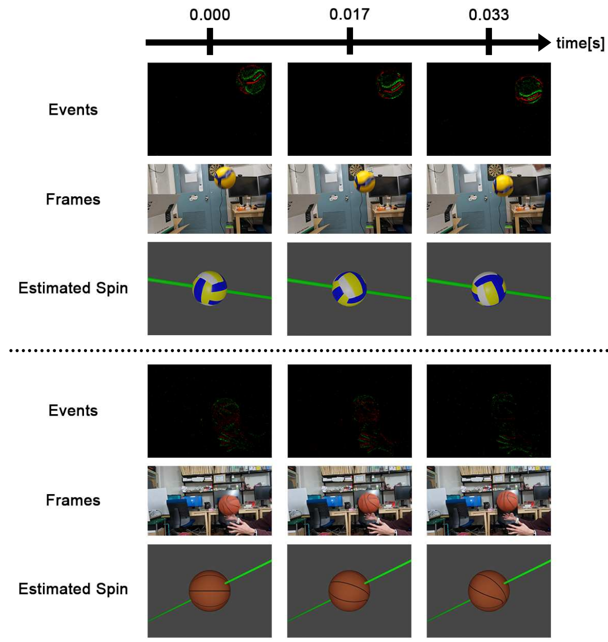

Figure 12. Qualitative evaluation on Lab Dataset. Estimated Spin images are rendered based on the estimated spin axis and angular velocity. The green line represents the spin axis.

5.4. Experiment on Real-world Dataset

Real-world Dataset. Real-world Dataset comprises sequences captured during a live volleyball match, serving to qualitatively verify the effectiveness of the proposed method in a practical setting. Images were captured from the spectator's seats using Prophesee Evaluation Kit 4 and a GoPro HERO 10 in a stereo configuration, as depicted in Figure 11. The initial position of the ball is provided through annotation.

Results. Figure 13 presents the visualization results of input events and estimated spins in Real-world Dataset. Estimating spin using image-based methods would be challenging due to severe motion blur when the ball is spiked, rendering the ball's texture indistinguishable. Conversely, the event camera captures the ball's texture clearly, without motion blur. Comparing the motion of the ball's texture captured by the event camera with the visualization results of the estimated spin, the directions of the spin axis appear to align closely. This consistency suggests that the proposed method is qualitatively effective in estimating spin.

6. Limitation and Future work

The proposed method necessitates not only events but also a bounding box of a ball. While, in our current experiment, the bounding box was manually provided, automating this annotation process is desirable due to its cost. Moreover, if translational motion is estimated simultaneously with ro-

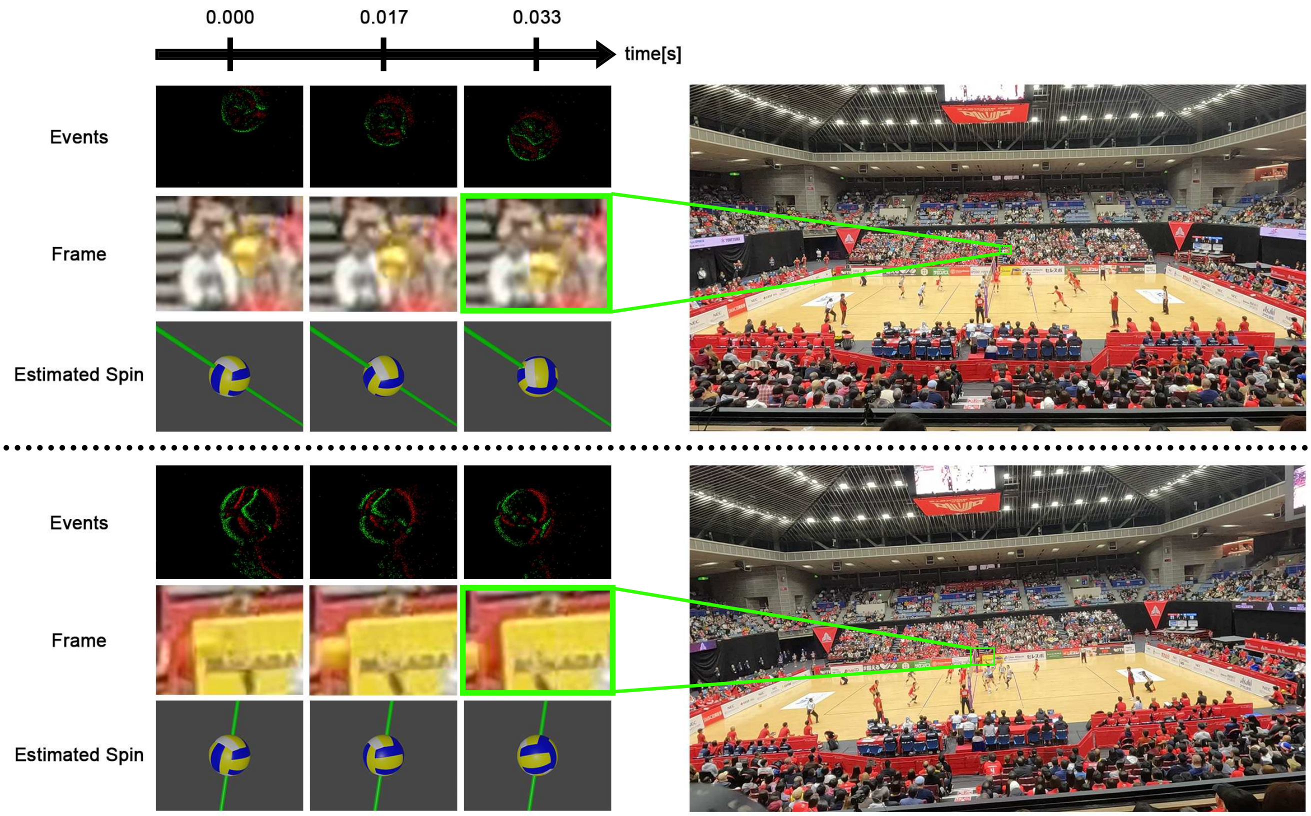

Figure 13. Qualitative evaluation on Real-world Dataset. Unlike frames with intense motion blur, the event clearly captures the texture of the ball.

tation, the number of parameters to be estimated increases and optimization becomes difficult. Hence, separating the process of ball position estimation from that of spin estimation and automating it would be beneficial. Several event-based ball position recognition methods [5, 9] have been proposed. Integrating these methods in future iterations could lead to the development of a more practical system.

Furthermore, due to the absence of ground truth data in Lab Dataset and Real-world Dataset experiments, we could only evaluate the performance of the proposed method qualitatively, rather than quantitatively. Moving forward, we aim to conduct similar experiments using, for instance, a ball embedded with a sensor. This will enable us to quantitatively evaluate the performance of the proposed method on real-world data while obtaining ground truth data.

7. Conclusion

This paper presents a novel method for ball spin estimation utilizing an event camera. Leveraging an event camera allows our method to effectively handle various lighting conditions and estimate fast ball spin. Moreover, unlike many existing methods, the proposed method does not depend on the type of ball and can be applied to balls used in various sports. The proposed approach estimates the normalized spin axis vector and angular velocity using events and the bounding box of the ball as inputs. Optimization with the Contrast Maximization framework is employed to estimate parameters from the events.

We conducted experiments to evaluate the performance of the proposed method using Synthetic Dataset, Lab Dataset, and Real-world Dataset. Quantitative experiments on Synthetic Dataset demonstrate that our method can estimate very fast spins more accurately than image-based methods. Also, experiments on Lab Dataset and Real-world Dataset qualitatively show that the proposed method is applicable to data captured by actual event cameras. These results exemplify the utility of event cameras in sports analysis.

Acknowledgement

This work was partially supported by JSPS KAKENHI, Grant Number JP23H03422.

References

[1] Peter Blank, Julian Hoßbach, Dominik Schuldhaus, and Bjoern M. Eskofier. Sensor-based stroke detection and stroke type classification in table tennis. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, page 93–100, New York, NY, USA, 2015. Association for Computing Machinery. 3

[2] Xiaopeng Chen, Ye Tian, Qiang Huang, Weimin Zhang, and Zhangguo Yu. Dynamic model based ball trajectory prediction for a robot ping-pong player. In *2010 IEEE International Conference on Robotics and Biomimetics*, pages 603–608, 2010. 2

[3] Sho Furuno, Kazuyuki Kobayashi, Tomoyuki Okubo, and Yosuke Kurihara. A study on spin-rate measurement using a uniquely marked moving ball. In *2009 ICCAS-SICE*, pages 3439–3442, 2009. 2

[4] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3867–3876, 2018. 2, 3

[5] Arren Glover and Chiara Bartolozzi. Event-driven ball detection and gaze fixation in clutter. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2203–2208, 2016. 8

[6] Thomas Gossard, Jonas Tebbe, Andreas Ziegler, and Andreas Zell. Spindoe: A ball spin estimation method for table tennis robot. *arXiv preprint arXiv:2303.03879*, 2023. 2

[7] Cheng Gu, Erik Learned-Miller, Daniel Sheldon, Guillermo Gallego, and Pia Bideau. The spatio-temporal poisson point process: A simple model for the alignment of event camera data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13495–13504, 2021. 3

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[9] Takuya Nakabayashi, Akimasa Kondo, Kyota Higa, Andreu Girbau, Shin'ichi Satoh, and Hideo Saito. Event-based high-speed ball detection in sports video. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports*, pages 55–62, 2023. 8

[10] Urbano Miguel Nunes and Yiannis Demiris. Entropy minimisation framework for event-based vision model estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 161–176. Springer, 2020. 3

[11] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Conference on robot learning*, pages 969–982. PMLR, 2018. 6

[12] Soichiro Sato and Masaki Aono. Leveraging human pose estimation model for stroke classification in table tennis. In *MediaEval*, 2020. 3

[13] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow. In *European Conference on Computer Vision (ECCV)*, pages 628–645, 2022. 3

[14] Hu Su, Zaojun Fang, De Xu, and Min Tan. Trajectory prediction of spinning ball based on fuzzy filtering and local modeling for robotic ping–pong player. *IEEE Transactions on Instrumentation and Measurement*, 62(11):2890–2900, 2013. 2

[15] Toru Tamaki, Haoming Wang, Bisser Raytchev, Kazufumi Kaneda, and Yukihiko Ushiyama. Estimating the spin of a table tennis ball using inverse compositional image alignment. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1457–1460, 2012. 4, 6

[16] Jonas Tebbe, Lukas Klamt, Yapeng Gao, and Andreas Zell. Spin detection in robotic table tennis. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 9694–9700. IEEE, 2020. 2

[17] Yifeng Zhang, Rong Xiong, Yongsheng Zhao, and Jianguo Wang. Real-time spin estimation of ping-pong ball using its natural brand. *IEEE Transactions on Instrumentation and Measurement*, 64(8):2280–2290, 2015. 2