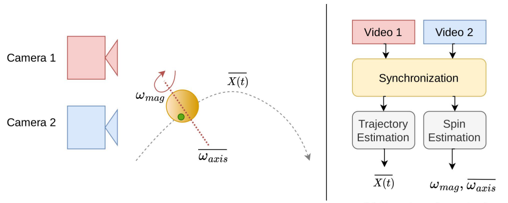
This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2025.3537116

Ball Trajectory and Spin Analysis from Asynchronous Videos

Aakanksha, Ashish Kumar, and Rajagopalan A. N.

Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India

Abstract—Existing systems for ball trajectory and spin estimation use embedded sensors or expensive high-frame-rate cameras which severely limits their accessibility. We propose an easy-to-setup low-cost vision sensor pipeline using two static asynchronous consumer-grade cameras. We also propose the use of epipolar geometry for synchronizing the cameras. We estimate 3D ball trajectory and spin with only one distinguishable feature on the ball. Mixture of Gaussians and adaptive color-based thresholding are used to localize the ball in 2D followed by triangulation. To estimate spin magnitude and axis, we employ feature detection and plane fitting. Extensive experiments with three different balls across multiple varied environments are reported and the approach is validated by arriving at the standard gravitational acceleration value from our estimated ball trajectory. For validating the spin, we compare our results with the true spin for a rotating ball fixed on a motor shaft. The average reprojection error was below 10 pixels for all our experiments and a maximum deviation of 17 RPM in spin magnitude was observed.



(a) Experimental Setup

(b) Overview of our pipeline

Index Terms—Ball detection, motion analysis, trajectory estimation, spin estimation, camera synchronization.

I. INTRODUCTION

Trajectory and spin estimation are fundamental tasks in ball sports [1]. Ball trajectory estimation involves finding the 3D location of the ball at any given time. Some sports have adopted using embedded Inertial Measurement Units (IMUs) to obtain ball trajectories [2] while others deploy proprietary setups comprising multiple expensive and sophisticated high-end vision sensors [4], [6]–[11]. In contrast, spin estimation is relatively under-explored [3], [4], [6], [12]–[14] despite being an integral part of advanced game-plays in most ball sports.

Using IMU enabled balls which are ball-specific or specialized sophisticated vision-sensor setups is often expensive which precludes their accessibility to individual players who could benefit from it during training and coaching sessions. We propose an easy-to-setup vision sensor pipeline for tracking a ball in 3D and for estimating 3D spin using two asynchronous consumer-grade cameras. Since the knowledge of corresponding 2D locations in at least two views at the same instant is a prerequisite for estimating 3D position, we propose a new vision-based approach for video synchronization as part of our pipeline.

For dynamic objects, there can be significant changes in scale and illumination during motion. Additionally, estimating spin for symmetric moving objects like a ball is highly challenging since it is difficult to detect the same distinguishing features across frames. Deep-learning (DL) based approaches for 2D object localization are data-dependent and do not generalize well across viewpoints and environmental conditions which are commonly encountered in end-user applications. Our approach can adaptively adjust to a new environment and ball, unlike DL methods which require a lot of data along with ground truth annotations for retraining. The proposed pipeline can be used to automatically generate annotations for DL models to aid with large-scale dataset creation. Our approach offers a potential low-cost alternative to the expensive Hawk-Eye [6] match officiating system. To summarize, our major contributions are -

1) We propose a pipeline to estimate ball trajectory and the 3D spin of a ball using a pair of unconstrained, asynchronous and static consumer-grade cameras.
2) Our approach estimates the ball centers in 2D in both the cameras using two-stage filtering. We use epipolar geometry to synchronize the captured videos and perform triangulation to locate the ball center in 3D. Feature detection and plane fitting is used for spin analysis.
3) We show results with multiple balls in different environments and validate our results on real data. Our method outperforms DL approaches especially in terms of generalizability across varying viewpoints and in detecting the moving ball.

II. PROPOSED METHODOLOGY

We want to estimate ball trajectory $\overline{X(t)} \in \mathbb{R}^3$, spin magnitude $\omega^{\text{est}}_{\text{mag}} \in \mathbb{R}^+$ and spin axis $\overline{\omega^{\text{est}}_{\text{axis}}} \in \mathbb{R}^3$ with $||\overline{\omega^{\text{est}}_{\text{axis}}}|| = 1$, given two cameras looking at the object of interest. Let C_{m_1} and C_{m_2} be calibrated cameras with projection matrices $\mathbf{P_1} = \mathbf{K_1}[\mathbf{R_1}|\overline{t_1}]$ and $\mathbf{P_2} = \mathbf{K_2}[\mathbf{R_2}|\overline{t_2}]$, respectively, placed some distance apart on tripods. The ball is being thrown from between them. Fig. 1 gives an overview of our proposed pipeline.

A. Ball Detection and Localization in 2D

The first step is to localize the ball in 2D. Let $\overline{I}_n$ be the undistorted n^{th} frame of a video taken by C_{m_1} or C_{m_2}. We use the pixel-based Mixture of Gaussians (MoG) model [15] to detect foreground pixels corresponding to the moving regions in each frame. This gives us a binary mask $\overline{F}_n = MoG(\overline{I}_n)$. But this is often noisy due to the motion of random objects and shadows. To address this issue, we use color-based thresholding at every spatial location x in $\overline{I}_n$ and get $F^{\text{color}}_n(x) = \begin{cases} 1, & \overline{T}_{min} \leq \overline{I_n}(x) \leq \overline{T}_{max} \\ 0, & \text{otherwise} \end{cases}$, where $\overline{T}_{min}$ and $\overline{T}_{max}$ are the minimum and maximum values of H, S and V in HSV space taken from a sampled region of interest (RoI) on the ball. $\overline{T}_{min}$, $\overline{T}_{max}$ and $\overline{I_n}(x)$ are all 3×1 vectors and these thresholds need to be adaptively selected only once for a new environment and setup. More details are provided in supplementary Sec. S1. The final ball mask

Corresponding author: Aakanksha (e-mail: aakankshajha30@gmail.com).

$F_n^B$is then obtained using $F_n^B = F_n \odot F_n^{\text{color}}$ where, $\odot$ is pixel-wise multiplication. We then fit a circular contour with maximum area on the derived mask F_n^B to localize the ball. The center of the circular contour, c_n, is our estimated ball center while the 2D radius r_n localizes the ball in the image plane.

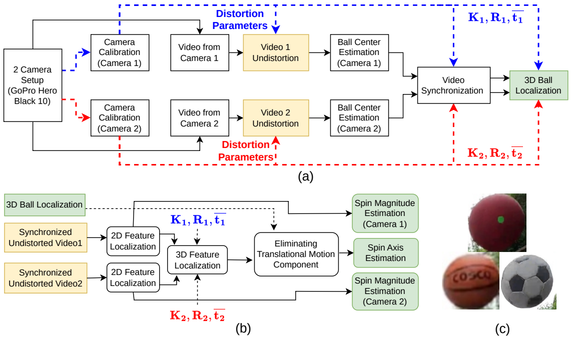

Distortion Parameters
$\mathbf{K_1}, \mathbf{R_1}, \overline{\mathbf{t_1}}$
2 Camera Setup (GoPro Hero Black 10)
Camera Calibration (Camera 1)
Video from Camera 1
Video 1 Undistortion
Ball Center Estimation (Camera 1)
Video Synchronization
3D Ball Localization
Camera Calibration (Camera 2)
Video from Camera 2
Video 2 Undistortion
Ball Center Estimation (Camera 2)
Distortion Parameters
$\mathbf{K_2}, \mathbf{R_2}, \overline{\mathbf{t_2}}$
(a)
3D Ball Localization
Synchronized Undistorted Video1
Synchronized Undistorted Video2
2D Feature Localization
2D Feature Localization
$\mathbf{K_1}, \mathbf{R_1}, \overline{\mathbf{t_1}}$
3D Feature Localization
$\mathbf{K_2}, \mathbf{R_2}, \overline{\mathbf{t_2}}$
Eliminating Translational Motion Component
Spin Magnitude Estimation (Camera 1)
Spin Axis Estimation
Spin Magnitude Estimation (Camera 2)
(b)
COSCO
(c)

Fig. 1: Overview of our proposed pipeline for (a) estimating 3D ball trajectory and (b) 3D spin. We experiment with the three balls depicted in (c) and report results for different environments in Table 2.

B. Synchronization

Due to the dynamics of the scene, the efficacy of our setup relies on careful time-synchronization of the cameras. We experiment with two approaches for synchronization - (a) Audio-based and (b) Audio-free. The audio-based approach for video synchronization (Approach I) uses a freely available software [16] to align corresponding video frames by comparing their audio similarities. In contrast, the proposed audio-free approach is grounded in epipolar geometry. For our approach, we first need to estimate the fundamental matrix, $\mathbf{F}$. Two variants exist based on how $\mathbf{F}$ is estimated. When sufficient features are present in the scene and SIFT features [17] are used along with the 7-point algorithm [18] to estimate $\mathbf{F}$, we denote it as Approach II A. Approach II B on the other hand uses camera parameters estimated during calibration to obtain $\mathbf{F}$ using the equation $\mathbf{F} = \mathbf{K_2}^{-T}[\overline{\mathbf{t}}]_{\mathbf{X}}\mathbf{R}\mathbf{K_1}^{-1}$. where, $\mathbf{K_1}$ and $\mathbf{K_2}$ are the camera intrinsic matrices, $[\overline{\mathbf{t}}]_{\mathbf{X}}$ is the skew-symmetric matrix representation of the translation between C_{m_1} and C_{m_2} which was derived to be $\overline{t} = \overline{t_2} - \mathbf{R_2}\mathbf{R_1}^T\overline{t_1}$. Here, $\mathbf{R_1}$ and $\mathbf{R_2}$ are the rotation matrices and $\overline{t_1}$ and $\overline{t_2}$ are the translation vectors for cameras C_{m_1} and C_{m_2} respectively. The rotation between the two cameras was derived to be $\mathbf{R} = \mathbf{R_2}\mathbf{R_1}^T$. Let the set of detected ball centers in videos V_{C_1} and V_{C_2} be B_1 and B_2 respectively. We then randomly select a center b_m^1 from V_{C_1} corresponding to the m^{th} frame and find its corresponding epipolar line in V_{C2} as $\overline{l} = \mathbf{F}b_m^1$. We then iterate over the set B_2 and the frame with detected ball center having minimum perpendicular distance from $\overline{l}$ is our corresponding frame in V_{C_2} for m^{th} frame in V_{C_1}. The synchronized videos are then used for 3D localization. Note that the ball center detected in multiple frames can be used for more robust synchronization. See Supplementary (Sec. S6) for details.

C. Localization in 3D

For 3D localization, we triangulate the detected ball centers. In particular, if frame i in C_{m_1} and frame j in C_{m_2} correspond to the same time instant, t, and c_i^1 and c_j^2 are the corresponding 2D ball centers with $\mathbf{P_1}$, $\mathbf{P_2}$ being the respective projection matrices, then, $\overline{c_i^1} = \mathbf{P_1}\overline{X}$ and $\overline{c_i^2} = \mathbf{P_2}\overline{X}$ where $\overline{X}$ is the ball center in 3D. This lets us write $\overline{c_i^1} \times \mathbf{P_1}\overline{X} = 0$ and $\overline{c_j^2} \times \mathbf{P_2}\overline{X} = 0$ where × denotes cross-product. Gauss-Newton optimization is then used to solve for $\overline{X(t)}$ with the initial estimate obtained using Direct Linear Transform (DLT) [19].

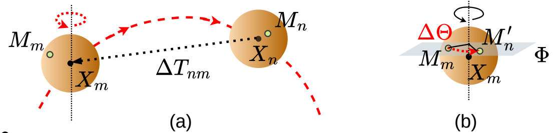


Fig. 2: The figure illustrates our approach of (a) translation motion component compensation to mimic pure rotational motion, and (b) plane and normal estimation through feature points on the rotating ball.

D. Spin Analysis

To estimate the spin magnitude, $\omega_{\text{mag}}^{\text{est}}$, we crop a bounding box of size $1.6r_n \times 1.6r_n$ around the detected 2D ball center c_n and perform feature detection within it for each frame n. Note that r_n is the radius of the ball in the image plane and the side length of bounding box is reduced empirically to $1.6r_n$ from $2r_n$ to avoid mis-detections due to the presence of background pixels within the bounding box. The detection of the feature in the set of all ordered frames is modeled as a square wave, $f_{\text{sq}}(n) = \begin{cases} 1, & \text{if feature is detected} \\ 0, & \text{otherwise} \end{cases}$. The Discrete Fourier Transform (DFT) of f_{sq} is taken and the frequency corresponding to the dominant magnitude in the spectrum is the estimated spin magnitude. Mathematically, $\omega_{\text{mag}}^{\text{est}} = \frac{F_s k_{\text{est}}}{M} \times 60$, where, k_{est} is the most dominant component in the DFT of f_{sq}, M is the length of f_{sq} which we consider as one or more complete cycles of the square wave, F_s is the camera frame rate and $\omega_{\text{mag}}^{\text{est}}$ is the estimated spin magnitude in rotations per minute (RPM).

To estimate the spin-axis of the rotating ball, we employ a 3D plane-fitting approach. Given one distinguishing feature on the ball, we first localize this feature in 2D and then estimate its 3D location using triangulation. Let I_m^1 and I_n^1 be frames at two different time instants for the same video, and $(\overline{M_m}, \overline{M_n})$ and $(\overline{X_m}, \overline{X_n})$ be the corresponding 3D locations of the feature centers and the centers of the ball, respectively. We obtain the 3D displacement of the ball center as $\Delta\overline{T_{nm}} = \overline{X_n} - \overline{X_m}$ between frames n and m. We compensate for the translation component between the two frames using $\overline{M_n^{'}} = \overline{M_n} - \Delta\overline{T_{nm}}$ where $M_n^{'}$ is the marker location in 3D for frame n with the translation component eliminated. This process has also been depicted in Fig. 2. Once we have nullified the translation components of at least three feature centers, we fit a plane, Φ, defined as $Z = aX + bY + c$ through the centers using least squares minimization. The corresponding unit normal $\overline{\omega_{\text{axis}}^{\text{est}}} = \pm[-a, -b, 1]/\sqrt{a^2 + b^2 + 1}$ gives the spin axis of the ball. Note that both the unit normal represent the same axis but in opposite directions. Additionally, while only one distinguishable feature is necessary, multiple distinct features on the ball can be used to improve robustness of spin estimation as shown in Sec. III.

III. EXPERIMENTS

We show the effectiveness of our approach on three different types of balls - (1) Ball A: a light-weight red coloured ball (2) Ball B: standard football (3) Ball C: a standard Size-5 basketball and use two GoPro Hero Black 10 cameras mounted on their standard tripods, in our setup. For Ball A, a small circular green marker is stuck on the ball to serve as a distinguishable feature while the manufacturer logo is used as the distinguishable feature for Ball B and Ball C. Videos are recorded at 120 fps and the resolution is kept at 1920 × 1080 for Ball A and at 3840 × 2160 for Ball B and Ball C. This is done to capture finer details of the logo for Balls B and C. The experiments are conducted in one indoor and two outdoor environments for each of the three balls. The indoor setup (E1) is an enclosed room with artificial lighting, and the first outdoor setup (E2) is a residential area with concrete structures. The second set of outdoor experiments is conducted in open areas, and we consider an open field for Ball A (E3), a football field for Ball B (E4), and a basketball court for Ball C (E5). This is done to closely mimic the environmental setups where our work has potential applications. To establish the robustness of our approach to different lighting conditions, we conduct experiments on Ball B and Ball C under bright sunny conditions (E4a, E5a) as well as on overcast cloudy days (E4b, E5b).

All the processing was performed on a system with Intel Xeon(R) CPU E5-1620 v3 @ 3.50GHz × 8. *MATLAB* was used for camera calibration with a 10×7 checkerboard having a square size of 34 mm. We fix our world coordinates to have the +Z axis parallel to the camera optical axis and +Y axis towards the ground. Radial distortion was removed for each video during pre-processing.

Synchronization was done using each of the three approaches detailed in Sec.II-B and the results are shown in Table 1. Ground truth is obtained by manually examining the alignment of frames corresponding to the instant of first ball bounce in both videos. We report the mean and standard deviation of error in the alignment of frames over three ball throws each for different types of balls, in indoor as well as outdoor settings. The overall errors are low for our proposed approaches (Approach II A and II B), with Approach II A giving the best results. This clearly shows the efficacy of our proposed approach in automatically aligning frames from asynchronous videos, even in the absence of any audio cues or special hardware.

Table 1: The mean and standard deviation (Std. Dev.) of frame differences (FD) from ground truth for different balls and environments.

		Approach I		Approach II A		Approach II B	
		Mean FD	Std. Dev. FD	Mean FD	Std. Dev. FD	Mean FD	Std. Dev. FD
Ball A	E1	0.33	0.58	1.33	0.58	0.67	0.58
	E2	0.33	0.58	1.33	0.58	1.33	1.15
	E3	1.00	1.00	0.33	0.58	0.83	0.58
Ball B	E1	0.00	0.00	0.67	0.58	0.67	0.58
	E2	1.00	1.00	1.67	1.15	12.00	6.08
	E4a	0.00	0.00	0.00	0.00	2.67	2.08
	E4b	0.33	0.58	0.33	0.58	4.00	1.73
Ball C	E1	8.00	13.00	0.33	0.58	1.67	0.58
	E2	1.00	1.00	2.33	0.58	3.67	1.15
	E5a	0.33	0.58	0.67	0.58	4.67	1.15
	E5b	0.33	0.58	0.33	0.58	4.00	1.00
Overall		1.15	1.72	0.85	0.58	3.29	1.52

A. Estimating Trajectory and Spin for Ball Throws

We report mean reprojection errors for the estimated trajectories as a metric to establish the quality of our estimated trajectories. Reprojection error for an estimated 3D point is defined as the Euclidean distance between projections of the estimated 3D point and the actual 3D point onto the image plane. Table 2 shows reprojection errors corresponding to trajectory estimation for Balls A, B and C for 3 ball throws each, in different environments. Care was taken to ensure variety in the spin axis, especially in terms of the dominant spin axis. A reprojection error of less than 10 pixels is observed on average for Ball A with consistently low point-wise standard deviation (under 6 pixels). For Ball B and Ball C, the reprojection errors are under 13 pixels each with the corresponding standard deviations under 10 and 9 pixels respectively. Note that while the reprojection error is still low for Ball B and Ball C, it is higher when compared to Ball A. We attribute this to the lower contrast of the colours of Ball B and Ball C with respect to the background. A similar trend is observed in point-wise standard deviation as well. Additionally, note that the reprojection errors are relatively higher in $E3$, $E4$ and $E5$ environments as compared to $E1$ and $E2$. We attribute this to the windy outdoor conditions and the lightweight nature of the tripods which results in small but perceptible movements of the camera. More details are provided in supplementary Sec. S5.

Table 2: The mean reprojection errors (Rep. Err.) and standard deviations (in pixels) are reported for the estimated 3D trajectories along with the estimated spin magnitude (in RPM) and spin axis. Here, "Ball A - E2 - 3" refers to the third ball throw in environment $E2$ and so on.

Video Number	Trajectory Estimation				Spin Estimation		
	Reprojection Error		Standard Deviation		Estimated RPM		Estimated Normal
	Cam 1	Cam 2	Cam 1	Cam 2	Cam 1	Cam 2	
Ball A - E1 - 1	2.39	2.52	1.35	1.42	635.29	635.29	[0.00, 0.99, 0.02]
Ball A - E1 - 2	3.10	3.54	2.16	2.59	553.84	553.84	[0.90, 0.06, 0.42]
Ball A - E1 - 3	3.34	3.51	2.03	2.12	600.00	600.00	[-0.20, 0.19, 0.96]
Ball A - E2 - 1	2.32	2.13	1.08	0.99	218.18	218.18	[-0.90, 0.42, 0.08]
Ball A - E2 - 2	2.36	2.13	1.27	1.15	288.00	288.00	[0.24, 0.84, 0.48]
Ball A - E2 - 3	4.83	4.83	2.18	2.01	327.27	327.27	[-0.99, 0.11, -0.01]
Ball A - E3 - 1	5.89	6.28	3.83	4.44	342.86	352.94	[-0.32, 0.90, -0.27]
Ball A - E3 - 2	8.69	8.36	1.01	1.35	310.34	338.03	[0.99, -0.06, -0.02]
Ball A - E3 - 3	8.58	9.57	4.73	5.59	363.64	363.64	[-0.77, -0.43, 0.46]
Ball A (Overall)	4.61	4.76	2.18	2.41	-	-	-
Ball B - E1 - 1	7.39	5.64	5.21	3.82	250.00	250.00	[-0.96, 0.24, -0.09]
Ball B - E1 - 2	7.37	5.56	4.76	3.63	428.57	461.53	[0.06, 0.98, 0.13]
Ball B - E1 - 3	12.4	9.40	9.20	6.88	462.09	500.00	[0.42, 0.89, -0.15]
Ball B - E2 - 1	4.10	3.90	2.56	2.44	248.27	252.63	[-0.99, -0.02, 0.01]
Ball B - E2 - 2	6.14	5.96	3.75	3.67	189.47	184.61	[0.02, 0.95, 0.31]
Ball B - E2 - 3	6.26	5.98	3.14	3.02	156.52	156.52	[-0.99, 0.07, -0.04]
Ball B - E4a - 1	9.83	9.75	4.44	4.41	230.77	230.77	[-0.95, 0.27, 0.16]
Ball B - E4a - 2	10.36	10.30	3.08	3.00	250.00	250.00	[-0.28, -0.93, 0.24]
Ball B - E4a - 3	10.33	10.19	5.04	4.98	260.87	240.00	[0.92, -0.38, -0.04]
Ball B - E4b - 1	10.46	10.33	5.22	5.05	285.71	285.71	[0.32, -0.91, 0.25]
Ball B - E4b - 2	9.72	9.62	6.41	6.27	240.00	272.72	[0.98, 0.08, 0.18]
Ball B - E4b - 3	8.60	8.64	4.34	4.26	181.81	176.47	[0.89, -0.33, -0.30]
Ball B (Overall)	8.58	7.94	4.76	4.29	-	-	-
Ball C - E1 - 1	7.67	7.70	5.78	5.75	230.77	230.77	[-0.99, 0.01, -0.04]
Ball C - E1 - 2	4.44	4.35	3.64	3.53	300.00	315.79	[0.99, 0.09, 0.11]
Ball C - E1 - 3	1.55	1.59	0.87	0.89	171.43	200.00	[0.62, 0.66, 0.43]
Ball C - E2 - 1	7.97	8.31	3.93	4.09	266.67	266.67	[0.87, 0.44, -0.19]
Ball C - E2 - 2	8.29	8.64	4.14	4.31	288.00	288.00	[-0.99, -0.02, -0.02]
Ball C - E2 - 3	11.48	12.09	7.19	7.60	313.04	300.00	[0.09, 0.37, -0.92]
Ball C - E5a - 1	8.83	9.86	7.03	8.22	260.87	260.87	[0.31, -0.95, -0.02]
Ball C - E5a - 2	7.77	7.26	4.95	4.80	193.54	195.65	[-0.95, 0.01, 0.32]
Ball C - E5a - 3	10.53	10.62	7.16	7.67	187.50	222.22	[0.96, 0.21, 0.19]
Ball C - E5b - 1	10.96	10.95	3.33	3.32	285.71	279.07	[0.94, 0.34, 0.01]
Ball C - E5b - 2	10.51	10.53	5.84	5.86	206.89	214.28	[-0.28, 0.92, 0.29]
Ball C - E5b - 3	8.97	8.78	3.78	3.69	214.28	240.00	[-0.13, 0.94, -0.32]
Ball C (Overall)	8.25	8.39	4.80	4.98	-	-	-

Table 2 also reports the spin estimation results for the same ball throws. Since our spin magnitude (RPM) estimation depends on only one camera, we individually estimate the RPM from both the cameras. As can be seen from the Table 2, results from both the cameras are consistent for each of the ball throws. The estimated spin axis is also consistent with what is observed visually in terms of the dominant spin axis.

To investigate the impact of using multiple distinct features in spin estimation, we experimented with Ball A by sticking three differently colored circular markers on it. The standard deviation decreased by 7.93 when using two markers as opposed to using only one marker for spin magnitude estimation while a decrease in average standard deviation from $[0.13, 0.07, 0.17]$ to $[0.03, 0.03, 0.05]$was observed for spin axis. This clearly establishes increased robustness when multiple features are used. Additional details and sample videos corresponding to Table 2 are included in supplementary Sec. S3.

B. Validation

Since it is non-trivial to obtain ground-truth values for the estimated trajectory as well as spin, we compare the results of our pipeline with measurable quantities in a controlled setup for validation. Without loss of generalizability, experiments were done only on Ball A.

Table 3: Ball A rotation validation results show consistency between the estimated (Est.) and ground truth (GT) values for both cameras (L and R)

RPM (GT)	RPM (Est. (L))	RPM (Est. (R))	Err. (RPM-L)	Err. (RPM-R)	Spin Axis (GT)	Spin Axis (Est.)	Err. (Axis)
120	125.69	120.00	5.69	0.00	[0.71, 0.00, 0.71]	[0.71, -0.03, 0.71]	[0.00, 0.03, 0.00]
180	186.82	189.21	6.82	9.21	[0.92, 0.00, 0.38]	[0.86, -0.19, 0.46]	[0.06, 0.19, 0.08]
100	109.72	104.13	9.72	4.13	[0.00, 1.00, 0.00]	[-0.01, 0.99, -0.08]	[0.01, 0.01, 0.08]
240	257.14	250.79	17.14	10.79	[0.38, 0.00, 0.92]	[0.32, 0.34, 0.88]	[0.06, 0.34, 0.04]
140	144.80	149.14	4.80	9.14	[1.00, 0.00 , 0.00]	[0.99, -0.01, -0.07]	[0.01, 0.01, 0.07]
Overall Error (RPM)			8.83	6.65	Overall Error (Axis)		[0.028, 0.116, 0.054]

For 3D localization, we validate using three quantities - (i) The reprojection error between the $2D$ ball centers and the triangulated 3D ball centers was calculated and this was found to be low as already reported in Table 2. (ii) Reprojection error was calculated between ten fixed and measured 3D points on the ground and wall surfaces and their triangulated counterparts, and the mean error value was found to be 1.8 pixels which is low as well. (iii) Gravitational acceleration was estimated by double differentiating the ball's 3D position with time for six ball trajectories and an average value of $9.18m/s^2$ with standard deviation 0.747 was obtained. Note that this is within $1m/s^2$ difference from the standard value of $9.8m/s^2$.

We validate our spin-estimation pipeline by mounting Ball A on a rotating motor with controllable speed. To verify the spin axis, we align the rotating motor shaft to different angles with respect to the world coordinates and compare the results as shown in Table 3. A maximum deviation of 17 RPM is observed while the major rotational axis remains consistent with the measured ground truth. Additional details are included in supplementary in Sec. S2.

C. Comparisons with Learning-based Approaches

We compared our proposed ball detection approach with three existing DL based approaches [20]–[23] for football detection to establish superior generalizability across views and scales. [20], [23] are trained on 20,000 frames and 11994 frames of broadcast videos of football matches. [21] is a standard YoLoV5 object detector trained on a curated dataset [24] containing 1340 football images captured in the wild while YoLoV9 [22] is trained on the MS-COCO [5] dataset. The quantitative results are reported in Table 4 for two videos for moving ball detection. Video 1 corresponds to the "*Ball A - E1 - 1*" video reported in Table 2, and Video 2 is a publicly available clip [1]. The numbers denote the percentage of frames where the ball was successfully localized using each of the approaches. A ball is considered successfully localized if the center of the fitted circular contour lies on the ball in frame. Our approach significantly outperforms both methods on the videos. While our approach is able to detect and localize the ball throughout, [21] can localize the ball only at closer range while [20], [23] localizes only at higher depths. We attribute this to the fact that at higher depths, the football occupies a very small area of the frame like broadcast videos which [20], [23] were trained on. On the other hand, the images used for training YoLoV5 have the football occupying a much larger area of the frame which results in poor performance at higher depths. While [22] detects the ball better, there is still scope for improvement. Qualitative results for Video 2 are included in supplementary.

Table 4: Quantitative comparisons for ball localization. The numbers denote the percent of frames where the ball was successfully detected.

	FootAndBall [20]	YoloV5 [21]	YoloV9 [22]	WASB[23]	Ours
Video 1	0.0	0.04	0.63	0.0	0.99
Video 2	0.18	0.39	0.47	0.06	0.89

IV. LIMITATIONS

Our approach uses MoG during ball localization which encodes motion information and may lead to issues when moving humans are visible in the frames wearing clothes of similar color. In such situations, DL methods which are much better at detecting humans than small balls, can be leveraged to mask out the moving humans and the ball can still be detected using our approach. See supplementary for additional information.

ACKNOWLEDGMENT

This work was supported in part by the Sports Science and Analytics Project (SP22231231CPETWOSSAHOC) from the Ministry of Education, Government of India. We are grateful to Mr. Avinash Gupta for his help in setting up the initial experiments and Prof. Mahesh Panchagnula for insightful discussions.

[1] Source:https://youtu.be/xHC8ozrKgRA

REFERENCES

[1] Zhao, Z., Chai, W., Hao, S., Hu, W., Wang, G., Cao, S., Song, M., Hwang, J. & Wang, G. A survey of deep learning in sports applications: Perception, comprehension, and decision. *ArXiv Preprint ArXiv:2307.03353*. (2023)

[2] Adidas reveals the first fifa world cup official match ball featuring connected ball technology. Adidas AG. [Online]. Available: https://preview.thenewsmarket.com/Previews/ADID/DocumentAssets/618225.docx

[3] Gossard, T., Tebbe, J., Ziegler, A. & Zell, A. SpinDOE: A Ball Spin Estimation Method for Table Tennis Robot. *2023 IEEE/RSJ International Conference On Intelligent Robots And Systems (IROS)*. pp. 5744-5750 (2023)

[4] Achterhold, J., Tobuschat, P., Ma, H., Büchler, D., Muehlebach, M. & Stueckler, J. Black-Box vs. Gray-Box: A Case Study on Learning Table Tennis Ball Trajectory Prediction with Spin and Impacts. *Proceedings Of The 5th Annual Learning For Dynamics And Control Conference (L4DC)*. **211** pp. 878-890 (2023,6), https://proceedings.mlr.press/v211/achterhold23a.html

[5] Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C. Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. pp. 740-755 (2014)

[6] P. Hawkins and D. Sherry, "Video processor systems for ball tracking in games," *World Wide Pat. WO01/41884A1, Int. Pat. Applic. A63B71/06*, vol. 14, 2001.

[7] B. T. Naik and M. F. Hashmi, "Lstm-bend: Predicting the trajectories of basketball," *IEEE Sensors Letters*, vol. 7, no. 4, pp. 1–4, 2023.

[8] Xiao, Q., Zaidi, Z. & Gombolay, M. Multi-Camera Asynchronous Ball Localization and Trajectory Prediction with Factor Graphs and Human Poses. *ArXiv Preprint ArXiv:2401.17185*. (2024)

[9] J. Ren, J. Orwell, G. A. Jones, and M. Xu, "A general framework for 3d soccer ball estimation and tracking," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 3. IEEE, 2004, pp. 1935–1938.

[10] B. Nobahar, M. Shoaran, and G. K. Khosroshahi, "Ball trajectory estimation and robot control to reach the ball using single camera," *Journal of Control*, vol. 14, no. 3, pp. 75–87, 2020.

[11] B. Chakraborty, "A trajectory-based ball detection and tracking system with applications to shooting angle and velocity estimation in basketball videos," in *2013 Annual IEEE India Conference (INDICON).*, IEEE, 2013, pp. 1–6.

[12] H. Shum and T. Komura, "Tracking the translational and rotational movement of the ball using high-speed camera movies," in *IEEE International Conference on Image Processing 2005*, vol. 3, 2005, pp. III–1084.

[13] Y. Zhang, R. Xiong, Y. Zhao, and J. Wang, "Real-time spin estimation of ping-pong ball using its natural brand," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2280–2290, 2015.

[14] P. Blank, B. H. Groh, and B. M. Eskofier, "Ball speed and spin estimation in table tennis using a racket-mounted inertial sensor," in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 2–9.

[15] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, 2004, pp. 28–31 Vol.2.

[16] Blackmagic Design, "DaVinci Resolve 18. (2022)", https://www.blackmagicdesign.com/products/davinciresolve/

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[18] R. Hartley, "Projective reconstruction and invariants from multiple images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 1036–1041, 1994.

[19] I. Sutherland, "Three-dimensional data input by tablet," *Proceedings of the IEEE*, vol. 62, no. 4, pp. 453–461, 1974.

[20] J. Komorowski, G. Kurzejamski, and G. Sarwas, "Footandball: Integrated player and ball detector," in *15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 01 2020, pp. 47–56.

[21] G. Jocher, "YOLOv5 by Ultralytics," May 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[22] Wang, C., Yeh, I. & Mark Liao, H. Yolov9: Learning what you want to learn using programmable gradient information. *European Conference On Computer Vision*. pp. 1-21 (2025)

[23] Tarashima, S., Haq, M., Wang, Y. & Tagawa, N. Widely Applicable Strong Baseline for Sports Ball Detection and Tracking. *BMVC*. (2023)

[24] Q. Imaging, "Soccerballdetector dataset," https://universe.roboflow.com/quince-imaging/soccerballdetector, jun 2023, visited on 2023-08-30. [Online]. Available: https://universe.roboflow.com/quince-imaging/soccerballdetector

[25] J. Calandre, R. Péteri, L. Mascarilla, and B. Tremblais, "Extraction and analysis of 3d kinematic parameters of table tennis ball from a single camera," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 9468–9475.