AA 273: Project Proposal

Matthew Lee
leemat@stanford.edu

Shiv Wadwani
swadwani@stanford.edu

Abstract—**Ball tracking is widely used in professional sports for accurate officiating and analytics, but existing systems often rely on expensive infrastructure, such as high-speed cameras or embedded sensors. In this project, we investigate the feasibility of low-cost ball tracking solutions using classical state estimation techniques. Focusing on a simplified 3D tennis ball bounce scenario, we develop a physics-based motion model and generate synthetic observations through a pinhole camera model. Estimation algorithms including the Kalman Filter, Extended Kalman Filter, Unscented Kalman Filter, and Particle Filter are implemented to infer ball position under various noise, occlusion, and camera configurations. Performance is evaluated using mean squared error, filter stability, and sensitivity to measurement conditions. This study aims to identify practical trade-offs in sensing and computation for accessible and reliable ball tracking, providing guidance for future deployment in resource-constrained environments.**

I. Introduction

Object tracking is a well-studied problem that has matured significantly in accuracy, performance, and robustness. In many sports, this technology has been applied prolifically, enhancing both player experience and officiating accuracy. High-fidelity sensors paired with robust detection, estimation, and filtering techniques have enabled ball tracking systems to deliver fairer line calls and reduce disputes in sports like tennis, volleyball, and soccer. In this project, we focus on reviewing state estimation methods for ball tracking and investigate how these techniques can be adapted and evaluated for low-cost, minimal-infrastructure setups—especially for use in recreational or resource-constrained environments.

II. Literature Review

Some sports like soccer and American football use a "Local Positioning System" (LPS) through the integration of ultra-wideband (UWB) sensors and inertial measurement units (IMUs) within the ball. Companies like Kinexon are industry leaders in this space, with partnerships across major leagues such as the NFL, MLS, and NBA [1]. In these systems, a network of UWB beacons placed at known locations around the field emits high-frequency impulses (500 Hz). The sensor in the ball measures the time difference of arrival (TDOA) of these signals to triangulate its position in real-time, typically achieving accuracies of around 10 cm. The onboard IMU supplements this by providing data on spin, acceleration, and impact force. This method of tracking ball location in the field was first presented here [2]. These systems offer high accuracy and low latency [3], but require expensive sensing technology and are only suitable for sports where the

presence of embedded electronics does not significantly alter the dynamics of the ball.

Sports like tennis and cricket use triangulation-based techniques to determine ball trajectory and location at specific events. For instance, in tennis, the position of the ball at impact must be accurately determined to decide whether it is in or out, which directly affects the outcome of a point. Hawk-Eye Innovations is a leader in this space [4]. Their system uses 10–12 high-speed, synchronized cameras positioned in fixed, calibrated locations around the court or pitch. The method for reconstructing ball location using this technology was first described by Pingali [5], where 2D images from each camera are used to reconstruct the 3D position of the ball through triangulation. These systems achieve an accuracy of 10 mm and have completely automated the process of line calling in nearly every major tennis tournament around the world. However, they are expensive, require specific calibration protocols, and require direct line of sight between the cameras and the ball, which is not always available. Kumar [6] provides another lower cost alternative with similar methodology to determine the position of the ball in the context of tennis but still requires multiple synchronized cameras.

Other applications in robotics, gaming, and film use motion capture systems like OptiTrack [7] and VICON [8]. These systems rely on an array of high-speed, high-resolution cameras along with reflective or active markers affixed to the object being tracked. The markers, often infrared (IR), allow for precise tracking of pose and trajectory through triangulation. While these systems offer extremely high accuracy and are widely used in controlled environments, they are often impractical for sports applications involving balls. The markers can alter the ball's dynamics, and may break or detach upon impact, conditions that are common in real play.

In many situations, recreational players or sports arenas that do not have the available infrastructure would benefit from ball tracking solutions that do not rely on expensive equipment or calibration procedures. In cases where lower resolution and frame rate cameras are used (such as a smartphone camera), or where sensor measurements may be intermittent due to occlusions or lighting conditions, estimation techniques can be used to determine the most likely position of the ball. Kamble [9] provides a detailed survey of a wide range of research in estimation techniques as they relate to the specific problem of ball position tracking, along with a synthesis of their results. The Kalman Filter in particular is reviewed as a potential solution to the estimation problem where a triangulation approach may not be available. Other research,

like [10], shows good results using a CNN for ball detection along with a Kalman Filter to determine the position of a golf ball in flight.

However in sports where discontinuities may occur (like during a bounce or hit, such as in tennis) the traditional Kalman Filter algorithm may struggle to produce accurate position estimations. Reid [11], Cox [12], and Genovese [13] have all formulated augmentations to the Kalman Filter that may address this problem for the given application. Reid and Cox detail a "Multiple Hypothesis Tracking" algorithm that is traditionally used to track the trajectories of multiple objects at once. Kittler [14] suggests this algorithm can also be adapted to track a single system with changing dynamics, like a bouncing tennis ball. Similarly, Genovese outlines an "Interacting Multiple Model" algorithm that directly tackles the problem of abruptly changing dynamics by smoothly transitioning between whichever dynamical model is most likely driving the system state at every iteration.

Such adaptations to the traditional Kalman Filter can also be extended to more advanced state estimation techniques like the Extended Kalman Filter, Unscented Kalman Filter, and Particle Filter. Particle Filters are especially useful in handling nonlinear dynamics and multimodal uncertainty, which arise in scenarios with discontinuities or occlusions. Yan [15] introduced a dual-model approach within the particle filter to separately model flight and impact phases. Cheng [16] further improves robustness by proposing an "automatic recovery" method that reinitializes the particle filter when tracking fails. Failure is detected by monitoring the degeneracy of particle weights. Instead of resampling, a new particle distribution is generated using 2D detections from synchronized multi-view images, reconstructed into 3D using homography, and filtered for consistency. This reinitialization effectively restores the belief state. Cheng also introduces an anti-occlusion model that adaptively excludes low-confidence camera views when computing measurement likelihoods. These enhancements significantly improve tracking accuracy in complex environments, increasing success rates from 63% to over 99%. However, the particle filter also has drawbacks due to its computational time complexity and particle impoverishment or degeneracy.

With the fast-paced advancement of deep learning, transformer architectures have also been applied to the problem of ball tracking, particularly for small, fast-moving objects. Yu [17] proposes "TrackFormer", a vision transformer-based model designed to address challenges such as motion blur, small object scale, and high-resolution input. The model improves spatial localization through global context modules and semantic enhancement. While this approach differs from traditional filtering-based methods, it offers a complementary data-driven alternative for estimating object trajectories from raw video. Their work also introduces LaTBT, the first large-scale dataset for tracking tiny balls in sports like badminton and squash, providing a valuable benchmark for evaluating future tracking algorithms.

III. Proposed Work

Current ball tracking solutions in sports often rely on expensive infrastructure involving high-speed cameras or embedded sensors. This project aims to explore the feasibility of lower-cost alternatives for tennis ball tracking using minimal sensing and classical state estimation methods. We focus on a simplified 3D bounce scenario as a baseline. The goal is to evaluate how well standard filtering techniques can estimate ball position under sparse and noisy observations.

A physics-based ground truth model will be used to simulate the ball's true trajectory. The dynamics will be modeled as:

$$x_{t+1} = x_t + \Delta t \cdot \dot{x}_t + w_{t,x}$$
$$y_{t+1} = y_t + \Delta t \cdot \dot{y}_t + w_{t,y}$$
$$z_{t+1} = z_t + \Delta t \cdot \dot{z}_t + \frac{1}{2} \cdot \Delta t^2 \cdot g + w_{t,z}$$

where g is gravitational acceleration and $w_t \sim \mathcal{N}(0, Q)$ represents process noise. The full state vector (3D position coordinates) is $p_t = [x, y, z]^\top$. For ground truth simulation, impact dynamics at bounce events will be handled using simplified models such as conservation of momentum, i.e.,

$$v_{t+1} = v_t + \Delta t \cdot \frac{1}{m} f_{\text{impact}}(p_t, v_t)$$

where f_{impact} can be derived from basic rigid-body contact models in physics literature. Note $p_t = [x, y, z]^T$ and $v_t = [\dot{x}, \dot{y}, \dot{z}]^T$.

Measurements will be generated synthetically using a pinhole camera model, where each camera projects the 3D world coordinates into 2D image coordinates. Cameras will be assumed to be in fixed, known locations, and measurement noise will be added to simulate real-world imperfections.

Estimation algorithms to be implemented include the Kalman Filter (KF), Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Particle Filter (PF). These filters will be tasked with estimating the ball's position from noisy, intermittent image-based observations. Performance will be evaluated using metrics such as mean squared error (MSE), filter divergence, and sensitivity to initialization. The effect of different factors such as number of cameras, measurement frequency, and initial state uncertainty will be systematically studied.

The goal of this study is to understand the trade-offs between algorithmic complexity, sensor setup, and tracking accuracy. By comparing these methods in a controlled setting, we aim to provide practical recommendations for the minimum sensing and algorithmic requirements needed to achieve reliable ball tracking in low-cost environments. A modular simulation and evaluation framework will be developed to support future experimentation.

References

[1] Kinexon. [Online]. Available: https://kinexon-sports.com

[2] A. Stelzer, K. Pourvoyeur and A. Fischer, "Concept and application of LPM - a novel 3-D local position measurement system," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 12, pp. 2664–2669, Dec. 2004, doi: 10.1109/TMTT.2004.838281.

[3] P. Blauberger, R. Marzilger, and M. Lames, "Validation of Player and Ball Tracking with a Local Positioning System," *Sensors*, vol. 21, no. 4, p. 1465, 2021.

[4] Hawk-Eye Innovations. [Online]. Available: https://www.hawkeyeinnovations.com/

[5] G. Pingali, A. Opalach and Y. Jean, "Ball tracking and virtual replays for innovative tennis broadcasts," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, Barcelona, Spain, 2000, pp. 152–156 vol.4, doi: 10.1109/ICPR.2000.902885.

[6] A. Kumar *et al.*, "3D Estimation and Visualization of Motion in a Multicamera Network for Sports," in *Proc. Irish Mach. Vis. Image Process. Conf. (IMVIP)*, Dublin, Ireland, 2011, pp. 15–19, doi: 10.1109/IMVIP.2011.12.

[7] OptiTrack. [Online]. Available: https://optitrack.com/

[8] Vicon. [Online]. Available: https://www.vicon.com/

[9] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, "Ball tracking in sports: a survey," *Artif. Intell. Rev.*, vol. 52, pp. 1655–1705, 2019, doi: 10.1007/s10462-017-9582-2.

[10] Y. Zhang, "Efficient Golf Ball Detection and Tracking Based on Convolutional Neural Networks and Kalman Filter," arXiv preprint arXiv:2012.09393, 2020. [Online]. Available: https://arxiv.org/abs/2012.09393

[11] D. Reid, "An algorithm for tracking multiple targets," in IEEE Transactions on Automatic Control, vol. 24, no. 6, pp. 843-854, December 1979, doi: 10.1109/TAC.1979.1102177.

[12] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 2, pp. 138-150, Feb. 1996, doi: 10.1109/34.481539.

[13] Genovese, A.F.. (2001). The interacting multiple model algorithm for accurate state estimation of maneuvering targets. Johns Hopkins APL Technical Digest (Applied Physics Laboratory). 22. 614-623.

[14] J. Kittler, W. J. Christmas, A. Kostin, F. Yan, I. Kolonias, and D. A. Windridge, "Memory architecture and contextual reasoning framework for cognitive vision," in *Scand. Conf. Image Anal.*, 2005, pp. 343–358.

[15] F. Yan, W. Christmas, and J. Kittler, "A tennis ball tracking algorithm for automatic annotation of tennis match," in *Proc. Br. Mach. Vis. Conf. (BMVC)*, 2005, pp. 619–628.

[16] X. Cheng, M. Honda, N. Ikoma, and T. Ikenaga, "Anti-occlusion observation model and automatic recovery for multi-view ball tracking in sports analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2016, pp. 1501–1505.

[17] J. Yu, Y. Liu, H. Wei, K. Xu, Y. Cao, and J. Li, "Towards Highly Effective Moving Tiny Ball Tracking via Vision Transformer," in *Adv. Intell. Comput. Technol. Appl., Lecture Notes in Comput. Sci.*, vol. 14864, Springer, Singapore, 2024, pp. 319–331, doi: 10.1007/978-981-97-5588-2_31.