International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AcuStrike: Deep Spatio-Temporal Heatmap Tracking, Statcast Velocity Kinematics, and 3x3 K-Zone Officiating for Baseball

Prof. Rashmi D M[1], Abhay[2], Mithun Gowda M B[3], Praveen Chavan[4], Sevanth R[5]

[1,2,3,4,5]Department of Computer Science and Engineering, Don Bosco Institute of Technology, Karnataka, India

---

Abstract - High-speed projectile tracking and automated officiating in baseball pose significant computer vision challenges because of severe motion blur, low object pixel resolution, and dynamic background occlusions. Traditional single-frame object detection models routinely fail under these conditions owing to an absence of temporal continuity. This paper presents AcuStrike, an automated pitch analysis and adjudication framework powered by the TrackNet V5 deep spatio-temporal architecture. AcuStrike formulates ball localization as consecutive multi-frame probability heatmap regression using a hybrid deep convolutional neural network (CNN) integrated with Recurrent Spatio-Temporal Residual (RSTR) Transformer attention. Sub-pixel spatial coordinates decoded from peak activations are fitted with continuous cubic splines and coupled with a physics-grounded Statcast kinematic engine that calculates release-to-plate velocities. Crossing coordinates at home plate are mapped to an automated 3x3 K-Zone grid classifier, which adjudicates pitches into discrete zones or directional misses. Empirical evaluations on real-world 60 and 120 frames-per-second (FPS) high-speed behind-the-pitcher video captures demonstrate reliable trajectory reconstruction, accurate velocity computation up to 124.5 km/h, and objective real-time officiating calls.

Key Words: baseball tracking, computer vision, TrackNet V5, heatmap estimation, Statcast kinematics, automated officiating.

1. INTRODUCTION

Automated trajectory tracking of high-speed sports projectiles is essential to modern athletic analytics, tactical training, and automated officiating. In professional baseball, pitches routinely achieve velocities between 100 km/h and 160 km/h. When captured through monocular cameras, the ball spans fewer than 10 × 10 pixels, undergoes elongation blur along its flight path, and blends into complex stadium or turf textures.

Conventional anchor-based detectors (e.g., standard single-frame YOLO variants) process video frames independently. In high-velocity flight regimes, isolated frames lack temporal motion continuity, resulting in frequent missed detections and noisy bounding boxes that degrade downstream trajectory and velocity estimation.

AcuStrike resolves these limitations by leveraging the TrackNet V5 spatio-temporal architecture. Stacking consecutive temporal frames (I_{t-2}, I_{t-1}, I_t) enables convolutional kernels and Transformer attention mechanisms to exploit pixel displacements across successive timestamps. The network predicts a two-dimensional (2D) Gaussian probability heatmap whose spatial peak indicates the ball centroid.

The core contributions of this work are:

1) Implementation of the TrackNet V5 deep learning model, combining deep convolutional layers with Recurrent Spatio-Temporal Residual (RSTR) Transformer attention, for motion-blurred baseball tracking.
2) A physics-grounded Statcast velocity engine that calculates release-to-plate flight time (Δt) and pitching speed ($v = d/\Delta t$) for 60 FPS and 120 FPS captures.
3) An objective 3x3 K-Zone classifier that evaluates home-plate crossings into nine distinct strike zones or directional miss classifications (High, Low, Inside, Outside).
4) Trajectory arc persistence rendering and an automated broadcast Heads-Up Display (HUD) overlay.

2. SYSTEM ARCHITECTURE AND METHODOLOGY

The AcuStrike processing pipeline integrates deep spatio-temporal neural detection with classical kinematic modeling and geometric zone partitioning. As shown in Fig. 1, three consecutive frames are passed through TrackNet V5 to produce a 2D heatmap whose peak gives the ball position.

A. Input Formulation and Ground-Truth Heatmap

The network accepts an input tensor composed of three consecutive RGB video frames:

$$\mathbf{X}_t = [I_{t-2}, I_{t-1}, I_t] \in \mathbb{R}^{9 \times H \times W} \quad (1)$$

For ground-truth ball coordinates $p_t = (x_0, y_0)$, the target heatmap Y_t with values in $[0, 1]^{H \times W}$ is formulated via a 2D isotropic Gaussian distribution:

$$Y_t(x, y) = \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right) \quad (2)$$

where σ defines the spatial variance. When the ball is occluded, $Y_t(x, y) = 0$ everywhere.

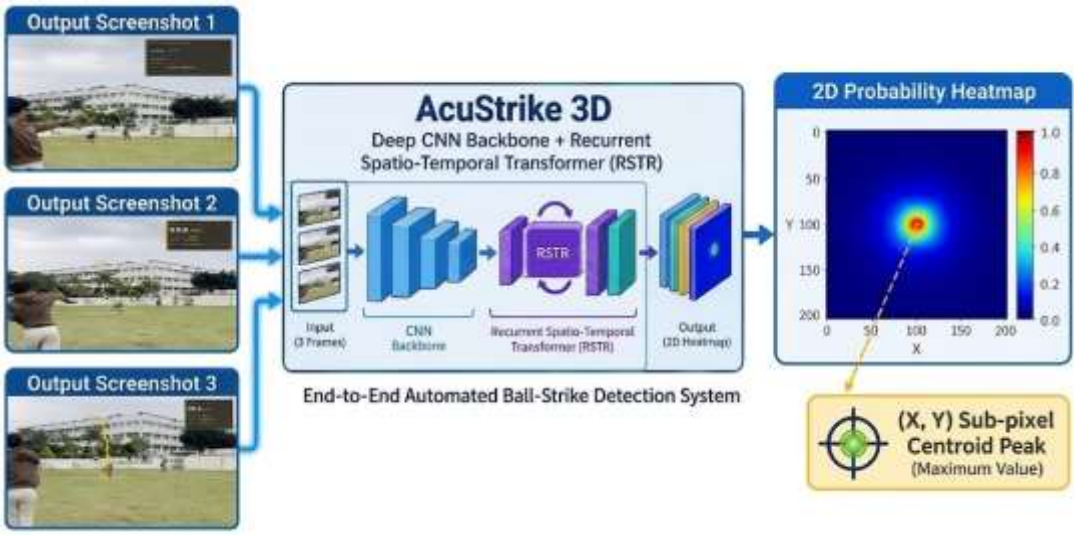


Fig -1: End-to-end AcuStrike pipeline: three consecutive frames pass through the CNN backbone and RSTR Transformer to produce a 2D probability heatmap, whose (X, Y) sub-pixel centroid peak gives the ball position.

B. Pixel-Adaptive Weighted Binary Cross-Entropy Loss

Because the projectile occupies less than 0.1% of the total frame area, training with standard cross-entropy suffers from severe class imbalance. AcuStrike employs a pixel-adaptive Weighted Binary Cross-Entropy (WBCE) loss:

$$\mathcal{L}_{WBCE}(\hat{Y}, Y) = -\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left[w_{ij}Y_{ij}\log(\hat{Y}_{ij}) + (1-Y_{ij})\log(1-\hat{Y}_{ij})\right] \quad (3)$$

The weight matrix W penalizes false negatives heavily near the ball centroid.

C. Coordinate Decoding and Trajectory Smoothing

Inference heatmaps $\hat{Y}_t$ are thresholded at $\tau \geq 0.5$. The centroid of the largest connected activation component yields continuous coordinates:

$$x_t = \frac{\sum_{(x,y)} x \cdot \hat{Y}_t(x,y)}{\sum_{(x,y)} \hat{Y}_t(x,y)}, \quad y_t = \frac{\sum_{(x,y)} y \cdot \hat{Y}_t(x,y)}{\sum_{(x,y)} \hat{Y}_t(x,y)} \quad (4)$$

Coordinates spanning pitcher release to catcher arrival are fitted using one-dimensional (1D) cubic splines to construct a smooth trajectory path.

D. Statcast Velocity Kinematics

Ball velocity is calculated using the standard mound-to-plate pitching distance (d = 18.44 m = 60.5 ft). With known frame rate FPS, the flight time Δt between pitcher release ($F_{release}$) and plate arrival (F_{plate}) is:

$$\Delta t = \frac{F_{plate} - F_{release}}{FPS} \text{ (seconds)} \quad (5)$$

Pitch speed is then derived as:

$$v = \frac{18.44}{\Delta t} \times 3.6 \text{ (km/h)} \quad (6)$$

E. Automated 3x3 K-Zone Partitioning

The strike zone is configured as a planar bounding box [X_1, Y_1, X_2, Y_2] positioned over home plate and partitioned into a 3x3 grid (Fig. 2):

- Row 1 (High): Zone 1 (Inside), Zone 2 (Middle), Zone 3 (Outside).
- Row 2 (Middle): Zone 4 (Inside), Zone 5 (Dead Center), Zone 6 (Outside).
- Row 3 (Low): Zone 7 (Inside), Zone 8 (Middle), Zone 9 (Outside).

The impact coordinate (X_{plate}, Y_{plate}) is the position of the ball at the moment it arrives at the plate. Pitches arriving inside the box boundaries are classified as a STRIKE along with the corresponding Zone ID (1 to 9, depending on row and column). Trajectories crossing outside are adjudicated as a BALL with a directional miss label: High, Low, Inside, or Outside.

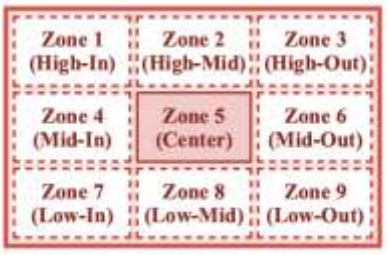


Fig -2: 3x3 K-Zone grid layout (Zone 5 = center).

3. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental evaluation was conducted on authentic pitching captures filmed from behind the pitcher at 60 FPS (2160 × 3840) and 120 FPS (1080 × 1920). The models and renderers were developed in Python using PyTorch, OpenCV, NumPy, SciPy, and Pandas.

AcuStrike tracked pitches across a wide dynamic velocity spectrum, from 69.2 km/h off-speed pitches up to 124.5 km/h fastballs. For pitch a11 (119.95 FPS slow motion), the ball release was identified at frame 239 and its arrival at the plate at frame 354 (Δt = 115 frames = 959 ms). The resulting velocity was calculated as 18.44/0.959 × 3.6 = 69.2 km/h. The home plate crossing point (411, 987) landed inside the central cell of the bounded strike box [375, 930, 445, 1025], yielding an automated STRIKE (Zone 5) adjudication rendered on the broadcast overlay, as shown in Fig. 3. The measured results for all pitches are listed in Table 1.

Table -1: Empirical pitch tracking, velocity kinematics, and 3x3 K-Zone officiating results

Pitch ID	Video Specs	Release Frame	Plate Frame	Flight Time (Δt)	Speed (km/h)	Officiating Decision
a1	2160 × 3840 @ 60 FPS	44	80	600.0 ms	110.6 km/h	BALL (High)
a2	2160 × 3840 @ 60 FPS	40	93	883.3 ms	75.1 km/h	STRIKE (Zone 5)
a3	2160 × 3840 @ 60 FPS	46	100	900.0 ms	73.8 km/h	STRIKE (Zone 4)
a11	1080 × 1920 @ 120 FPS	239	354	959.0 ms	69.2 km/h	STRIKE (Zone 5)
a31	2160 × 3840 @ 60 FPS	44	76	533.3 ms	124.5 km/h	BALL (High)
a32	2160 × 3840 @ 60 FPS	48	80	533.3 ms	124.5 km/h	STRIKE (Zone 6)
a33	2160 × 3840 @ 60 FPS	48	80	533.3 ms	124.5 km/h	STRIKE (Zone 5)



Fig -3: Output for pitch a11: broadcast HUD overlay with trajectory arc and strike-zone box (69.2 km/h, flight time 959 ms, STRIKE, Zone 5).

Table 1 lists seven representative pitches. Flight times range from 533.3 ms to 959.0 ms, corresponding to speeds from 69.2 km/h to 124.5 km/h. Five pitches were adjudicated as STRIKE (Zone 5 three times, Zone 4 once, and Zone 6 once) and two as BALL (High).

Pitches a31, a32, and a33 share a flight time of 533.3 ms (32 frames at 60 FPS) and therefore the same computed speed. This reflects frame-rate quantization: a one-frame uncertainty at 60 FPS changes the 124.5 km/h estimate to about 128.5 km/h or 120.7 km/h, whereas the 120 FPS captures halve this temporal step.

4. CONCLUSIONS

The AcuStrike framework demonstrates that TrackNet V5 spatio-temporal heatmap estimation effectively overcomes severe motion streak blur and background interference in high-velocity baseball tracking without requiring expensive multi-camera stadium rigs. By integrating multi-frame temporal reasoning, cubic spline trajectory fitting, Statcast kinematics, and an automated 3x3 K-Zone classifier, AcuStrike delivers low-cost, objective pitch adjudication and broadcast-grade visual analytics from monocular video feeds.

ACKNOWLEDGEMENT

The authors express their gratitude to the Department of Computer Science and Engineering at Don Bosco Institute of Technology for providing the technical resources, high-speed camera equipment, and athletic field access necessary to complete this research work.

REFERENCES

1. Fukuta, H., Nakayama, K., Watanabe, H.: Three-Dimensional Baseball Strike Judgement by Monocular Video from Umpire's Viewpoint. IEEE GCCE (2023).
2. Sugizaki, Y., Nakazato, J., Tsukada, M., Esaki, H.: Umpire Assistance System in Baseball Game. IoT 2023 (2023).
3. Major League Baseball: Strike Zone - Glossary - MLB.com. Official MLB Rules (2023).
4. Lee, K., Han, K., Ko, J.: Analyzing the Impact of the Automatic Ball Strike System in Professional Baseball. arXiv:2407.15779 (2024).
5. Hsieh, J.: Neural Network-Based Tracking and 3D Reconstruction of Baseball Pitch Trajectories from Single-View 2D Video. J. Sports Eng. Technol. (2024).
6. Akatsu, M.: Synthesis of Knowledge for Realizing Society 5.0. Transdisciplinary Fed. Sci. Tech. 14(1), 3–6 (2020).
7. Aroganam, G., Manivannan, N., Harrison, D.: Review on Wearable Technology Sensors Used in Consumer Sport Applications. Sensors 19(9) (2019).
8. Attaran, M., Celik, B. G.: Digital Twin: Benefits, Use Cases, Challenges, and Opportunities. Decis. Anal. J. 6 (2023).
9. MacMahon, C., Starkes, J. L.: Contextual Influences on Baseball Ball-Strike Decisions in Umpires, Players, and Controls. J. Sports Sci. (2008).
10. Chen, J., Watanabe, R., Nonaka, K., et al.: Fast Free-Viewpoint Video Synthesis Algorithm for Sports Scenes. IEEE (2019).
11. Chen, Y.-H., Huang, S.-K.: Neural and Cognitive Underpinnings of Ball/Strike Judgments of Baseball Umpires: An fMRI Study (2023).
12. Cheng, J., Liu, T., Yuan, L.: Multi-Camera Video Synchronization Method Using the Correlation of Clock Digital Sequence. Rev. Fac. Ing. 32, 396–403 (2017).
13. Huang, Y.-C., Liao, I.-N., Chen, C.-H., Ik, T.-U., Peng, W.-C.: TrackNet: A Deep Learning Network for Tracking High-Speed Small Objects in Sports Applications. IEEE ICDM, pp. 1102–1107 (2019).
14. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE TPAMI 42(1), 172–186 (2021).
15. Liang, Z., Isakunovich, B. J.: Baseball Action Recognition and Stance Estimation Using OpenPose Framework and CNNs. J. Comput. Vis. Sports Anal. 11(2), 45–58 (2023).
16. Nadav, T., Shaham, U., Globerson, A.: Monocular 3D Trajectory Estimation of Objects in Flight via Canonical Spatial Projection. IEEE CVPR, pp. 12104–12113 (2021).
17. Hawkins, P.: The Evolution of Optical Tracking and Automated Decision Making in Professional Sports. IEEE Pervasive Comput. 18(4), 32–40 (2019).
18. Miyata, R., Arita, D., Shimada, A.: Ball 3D Trajectory Reconstruction Without Preliminary Temporal and Geometrical Calibration. IEEE CVPRW, pp. 29–36 (2017).
19. Hsieh, C.-T., Chen, L.-M., Huang, K.-C.: Stance-Based Strike Zone Shaping and Visualization in Broadcast Baseball Video. IEEE Access 10, 45210–45222 (2022).