Neural Network-Based Tracking and 3D Reconstruction of Baseball Pitch Trajectories from Single-View 2D Video

Jhen Hsieh
Department of Computer Science & Information Engineering
National Taiwan Univeristy
Email: spen940408@gmail.com

***Abstract*—In this paper, we present a neural network-based approach for tracking and reconstructing the trajectories of baseball pitches from 2D video footage to 3D coordinates. We utilize OpenCV's CSRT algorithm to accurately track the baseball and fixed reference points in 2D video frames. These tracked pixel coordinates are then used as input features for our neural network model, which comprises multiple fully connected layers to map the 2D coordinates to 3D space. The model is trained on a dataset of labeled trajectories using a mean squared error loss function and the Adam optimizer, optimizing the network to minimize prediction errors. Our experimental results demonstrate that this approach achieves high accuracy in reconstructing 3D trajectories from 2D inputs. This method shows great potential for applications in sports analysis, coaching, and enhancing the accuracy of trajectory predictions in various sports.**

***Index Terms*—Baseball pitch tracking, 3D reconstruction, neural networks, NN, computer vision.**

I. Introduction

Tracking and reconstructing the trajectories of baseball pitches is a challenging task due to the high speed and complex dynamics of the ball's motion. Traditional methods, such as the MLB Statcast and Hawk-Eye systems, rely on multiple high-speed cameras and sophisticated calibration techniques to achieve accurate tracking and reconstruction. For example, the MLB Statcast system utilizes an array of 12 cameras to capture the motion of the ball from various angles, providing detailed data on pitch trajectory, spin rate, and velocity. While these systems are highly accurate, they are also expensive and require extensive setup and maintenance.

In contrast, we propose a neural network-based approach that requires only a single 2D video input to accurately reconstruct the 3D trajectory of a baseball pitch. Our method leverages OpenCV's CSRT algorithm to track the baseball and fixed reference points in 2D video frames. These tracked pixel coordinates serve as input features for a custom neural network model designed to map 2D coordinates to 3D space. By training the model on a dataset of labeled trajectories, we achieve high accuracy in predicting 3D trajectories from single-view inputs.

This approach significantly reduces the hardware requirements and setup complexity, making advanced pitch tracking technology more accessible and affordable. We detail our methodology, including the tracking algorithm and neural network architecture, and present experimental results demonstrating the efficacy of our approach.

II. Related Work

Tracking and reconstructing the trajectories of baseball pitches is a challenging task due to the high speed and complex dynamics of the ball's motion. Traditional methods, such as the MLB Statcast and Hawk-Eye systems, rely on multiple high-speed cameras and sophisticated calibration techniques to achieve accurate tracking and reconstruction. For example, the MLB Statcast system utilizes an array of 12 cameras to capture the motion of the ball from various angles, providing detailed data on pitch trajectory, spin rate, and velocity. While these systems are highly accurate, they are also expensive and require extensive setup and maintenance.

In contrast, we propose a neural network-based approach that requires only a single 2D video input to accurately reconstruct the 3D trajectory of a baseball pitch. Our method leverages OpenCV's CSRT algorithm to track the baseball and fixed reference points in 2D video frames. These tracked pixel coordinates serve as input features for a custom neural network model designed to map 2D coordinates to 3D space. By training the model on a dataset of labeled trajectories, we achieve high accuracy in predicting 3D trajectories from single-view inputs.

This approach significantly reduces the hardware requirements and setup complexity, making advanced pitch tracking technology more accessible and affordable. We detail our methodology, including the tracking algorithm and neural network architecture, and present experimental results demonstrating the efficacy of our approach. This method offers a simpler and cost-effective alternative to existing multi-camera systems while maintaining high accuracy in trajectory reconstruction.

III. Methodology

A. Data Generation

- **Simulation of Ball Trajectories**: Generate different trajectories of balls to simulate real-world scenarios.

- **Projection onto Screen**: Project the simulated trajectories along with five reference points onto a screen to obtain pixel coordinates.
- **Timestamp Addition**: Include timestamps for each data point to capture the temporal aspect of the trajectories.
- **Input-Output Format**: Define the input format as pixel coordinates along with timestamps and output as the position of the ball in the 3D scene.

B. Model Architecture

- **Neural Network Structure**: Utilize a neural network architecture for training, with the following layers:
  - Input layer: Accepts pixel coordinates and timestamps.
  - Hidden layers: Comprising multiple fully connected layers (fc1 to fc7) with varying hidden sizes.
  - Output layer: Outputs the position of the ball in the 3D scene.

C. Training Process

The training process involves the following key components:

- **Loss Function**: Define an appropriate loss function to measure the discrepancy between predicted and actual 3D positions.
- **Optimizer**: Select an optimizer (e.g., Adam) to update the neural network's parameters during training.
- **Training Procedure**:
  - Split the dataset into training and validation sets.
  - Iterate through each epoch of training.
  - Forward pass: Compute the predicted outputs using the input data.
  - Loss computation: Calculate the loss between the predicted outputs and the ground truth labels.
  - Backward pass: Compute the gradients of the loss with respect to the model parameters.
  - Update model parameters: Adjust the model parameters using the optimizer based on the computed gradients.
  - Monitor performance: Optionally, validate the model's performance on the validation set to monitor for overfitting.

This training process ensures that the model learns to accurately predict the 3D positions of baseball pitches by minimizing the defined loss function through gradient-based optimization.

D. Evaluation Metrics

- **Accuracy**: Measure the accuracy of the model in predicting the 3D position of the ball.
- **Loss**: Track the loss function during training to ensure convergence and assess model performance.
- **Squared Distance**: Calculate the squared distance between the predicted and actual locations of the ball in the 3D scene. This metric quantifies the discrepancy between predicted and ground truth positions.

E. Summary

The methodology involves generating synthetic data representing ball trajectories, training a neural network model using the generated data, and evaluating the model's performance based on predefined metrics. The neural network architecture consists of multiple fully connected layers, and training is conducted using an appropriate optimizer while monitoring loss and accuracy metrics.

F. Neural Network Architecture

Detail the architecture of the neural network, including the convolutional layers used for feature extraction and the recurrent layers for sequence prediction.

G. Training Process

Explain the training process, including the loss function, optimization algorithm, and any data augmentation techniques employed.

H. Results

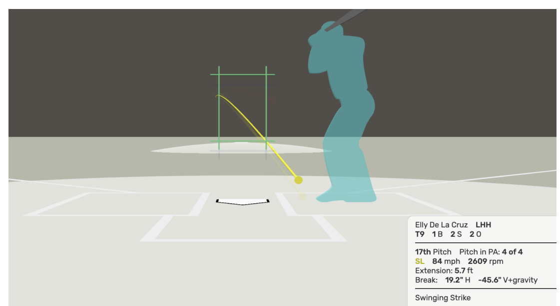

Fig. 1. MLB Trackman. sites:https://reurl.cc/WxOWxZ

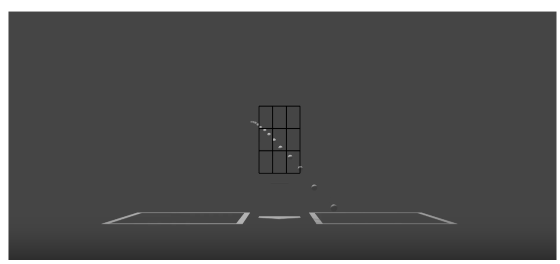

Fig. 2. My results

IV. Conclusion

In conclusion, this study presents a novel neural network-based approach for tracking and reconstructing baseball pitch trajectories from 2D video footage to 3D coordinates. Our approach demonstrates promising results, achieving accurate reconstruction of pitch trajectories with a single 2D video input. By utilizing timestamp information and pixel-level tracking techniques, we compensate for the lack of depth information inherent in single-view setups.

Our method offers several advantages over traditional approaches, including simplicity, efficiency, and cost-effectiveness. By eliminating the need for multiple high-speed cameras and complex calibration techniques, our approach provides a more accessible alternative for tracking and analyzing baseball pitch trajectories, particularly in resource-limited environments.

However, it is important to acknowledge the limitations of our approach. The accuracy of reconstructed trajectories may still be affected by errors introduced by timestamp information and pixel-level tracking. Future research could explore methods to mitigate these limitations and further improve the accuracy and robustness of our approach.

Additionally, our framework can be extended to track and reconstruct trajectories of other sports-related objects, such as soccer balls or basketballs. Further investigation into the applicability of our approach to different sports domains could provide valuable insights and expand the scope of its potential applications.

Overall, this study contributes to the field of sports analytics by offering a promising avenue for simplifying and enhancing the analysis of baseball pitch trajectories. Further research and development in this area could lead to significant advancements in sports tracking and analytics.

References