Project Yoriichi: Tracking and Control for Human-Robot Fencing via Vision Based Estimation

Abhishek Mathur, Deepam Ameria, Ishita Gupta, Kailash Jagadeesh

Abstract— **This paper presents Project Yoriichi, a vision-guided robotic fencing system that enables real-time human-robot interaction through responsive motion planning and control. Using a Franka Emika Panda robotic arm and an external RGB-D camera, the system tracks a brightly colored PVC pipe held by a human opponent through HSV-based color segmentation. The extracted 3D centroid of the opponent's baton is transformed into the robot's frame using extrinsic calibration and used to generate responsive trajectories that position the robot's baton in the anticipated path of the attack. A finite state machine manages transitions between passive observation and active defense, triggering trajectory generation based on spatial thresholds near a virtual defense plane. The robot mirrors the opponent's motion in real time, maintaining a stable stance while countering dynamic incoming trajectories.**

This work demonstrates an effective integration of low-cost perception and reactive control for competitive human-robot fencing scenarios. Future extensions will explore dynamic end-effector orientation and contact-aware trajectory generation to further enhance realism and responsiveness in interactive robotic defense.

I. INTRODUCTION

Human-robot interaction (HRI) has seen significant advancements in recent years, specifically from the context of motion tracking. It is becoming increasingly favourable for a vast number of real world applications particularly in dynamic and competitive environments such as sports and physical training. Among these, robotic fencing presents a unique challenge due to its high-speed motion, unpredictability, and the necessity for real-time decision-making. The development of robotic systems capable of engaging in fencing with a human requires precise motion tracking for imitating human operators, trajectory planning for responding to rapid incoming motions, and rapid reaction mechanisms to effectively defend against incoming attacks. Moreover, fencing uses sword-like tools which can be mounted easily on end effectors of robot arms to enable them as tracking agents.

Fencing, a sport that combines physical agility with strategic decision-making, has seen significant advancements in training and performance enhancement through the integration of advanced motion technologies. Motion capture systems, for instance, have been instrumental in analyzing and refining fencing techniques by providing precise measurements of body movements, thereby enhancing performance and reducing injury risks[1]. The use of wearable devices and virtual reality has further expanded the scope of training by offering real-time physiological data and simulated environments for mental conditioning[2].

In robotics, collaborative robots like the Franka Emika Panda have been increasingly utilized in research settings due to their versatility and ease of programming. Modular control stacks such as Franka-Interface and FrankaPy enable customizable and accessible interfaces for the Franka arm, facilitating rapid prototyping of new control methods[2]. Moreover, motion capture technology plays a crucial role in robotics by ensuring safe and effective human-robot interaction, as demonstrated by Vicon's high-precision tracking systems.[3]

In this paper, we discuss about the robotic fencing system using a Franka Emika robotic arm with low-cost motion tracking system developed for the project course of Robot Autonomy at Carnegie Mellon University. By integrating high-speed pose estimation with predictive motion planning, the system is designed to effectively recognize and defend against incoming attacks while maintaining safety constraints. The proposed approach contributes to the broader field of robotic sports training and HRI by demonstrating an adaptive and interactive robotic fencer, forming a smooth baseline for building imitation learning models for human motion in 3D.

II. RELATED WORK

Human-robot interaction in dynamic and adversarial tasks such as sports has garnered increasing attention within the robotics community. Among these, robotic systems that imitate or respond to high-speed human actions, such as sword fighting or fencing, pose unique challenges due to their requirements for rapid perception, real-time planning, and safe yet responsive actuation.

Namiki and Takahashi [3] present one of the earliest efforts toward developing a sword-fighting robot capable of reacting to human opponents. Their work demonstrates the effectiveness of using high-speed vision sensors for early detection of opponent motion and generating preemptive reactive strategies. While effective, their approach primarily focuses on deterministic motion templates and lacks the flexibility to generalize across variable trajectories.

In the domain of robotic manipulation, particularly using collaborative robots such as the Franka Emika Panda, significant progress has been made in developing modular control interfaces. Zhang et al. [2] introduce Franka-Interface and FrankaPy—open-source software stacks that allow rapid prototyping and real-time control of the Franka arm. These frameworks support safe and compliant motion generation, making them well-suited for interactive HRI scenarios like fencing.

In the domain of robotic manipulation, particularly using collaborative robots such as the Franka Emika Panda, significant progress has been made in developing modular control interfaces. Zhang et al. [2] introduce Franka-Interface and FrankaPy—open-source software stacks that allow rapid prototyping and real-time control of the Franka arm. These frameworks support safe and compliant motion generation, making them well-suited for interactive HRI scenarios like fencing.

Mirroring strategies for cobots have also been explored in tasks requiring intuitive, real-time collaboration. Ghadirzadeh et al. [5] propose a policy learning framework in which a robot learns to mirror human motion via deep reinforcement learning, allowing for adaptive and generalized responses in HRI contexts. Their findings are particularly relevant for our mirroring-based fencing defense strategy.

Similarly, the work by Rahmatizadeh et al. [6] explores vision-based robot imitation learning from demonstration videos, demonstrating how end-to-end deep models can infer and reproduce complex manipulation tasks. Although their domain is object manipulation, the principle of leveraging visual pose estimation for robotic motion imitation aligns closely with our objective of real-time defensive response.

Motion capture systems continue to be a gold standard for accurate pose estimation in human-robot systems. Vicon's tracking platforms [3] are widely adopted for their precision and low latency. However, due to their cost and infrastructure requirements, our approach instead investigates low-cost RGB-D tracking with HSV-based segmentation, which offers flexibility and accessibility in dynamic environments.

Filip [1] emphasizes the role of motion technology in enhancing athlete training, highlighting how real-time feedback and motion tracking improve performance and reduce injury—principles we incorporate in our fencing robot to simulate interactive and immersive training scenarios.

Our approach builds on these foundations by combining real-time low-cost vision tracking, dynamic mirroring strategies, and trajectory planning, thereby offering a reactive fencing robot platform for safe and adaptive human-robot engagement in adversarial sports-like environments.

III. METHODOLOGY

In robotic fencing, the forward process involves three key steps: perception, planning, and execution. First, the robot uses real-time sensing, using a Realsense camera placed in the ground frame, to track the opponent's baton movements. It then processes this information to predict the 3D pose and location of the incoming baton and determine the best response to the action. A finite state machine (FSM) helps the robot decide whether to defend/imitate or stay stagnated. Based on the predicted pose, the reverse process evaluates how the arm should contradict the path of the incoming baton by using motion imitation mirrored about the camera plane while maintaining a state aware pose and an on-arm camera calibrating the pose for the arm frame to apply the transforms from camera to robot world frame. The planner evaluates this in realtime for the desired end effector pose and plans the trajectory for the baton attached to the end effector to be able to reach the incoming baton. Success will be determined by the end effector baton always mimicking the human operator and countering the achieving perpendicular pose interacting at the middle of the approaching baton in any given incoming orientation which the user can create in 3D. Speed and enhanced variability for the final joint to change angle based on incoming angle will be a stretch goal. We will elaborate the current progress and our approach to the problem in the sections below.

Fig. 1. Mount to hold the pvc pipe

Fig. 2. Assembled Hardware System

A. Hardware Setup

We develop a mount for attaching the fencing baton (PVC baton) to the franka arm's end-effector. Figures 1 and 2 show the mount and the assembled system. We mount the camera in front of the robot base for maximum field of view for the human operator.

B. Perception

The perception subsystem is primarily dedicated to the real-time tracking of the movements of the human opponent, with a particular emphasis on estimating the pose of a PVC pipe - used here as a proxy for a fencing sword to simulate an attack scenario. In the initial stages of system design, we explored the feasibility of leveraging the Intel RealSense D435i camera that is already integrated and mounted on the end-effector of the robot arm. While this configuration provided some advantages in terms of proximity to the interaction point, further analysis of the system's operational requirements revealed that a static, externally mounted camera would yield a more stable, unobstructed, and consistent

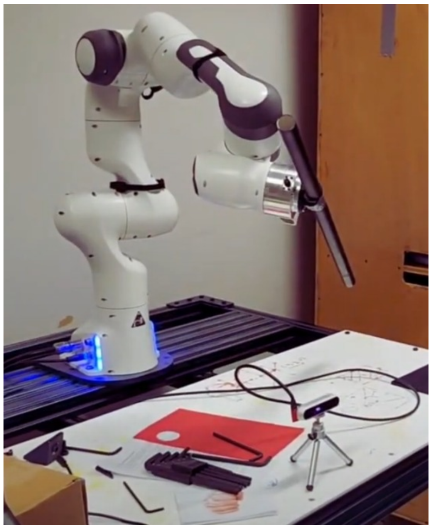

Fig. 3. Full System Setup

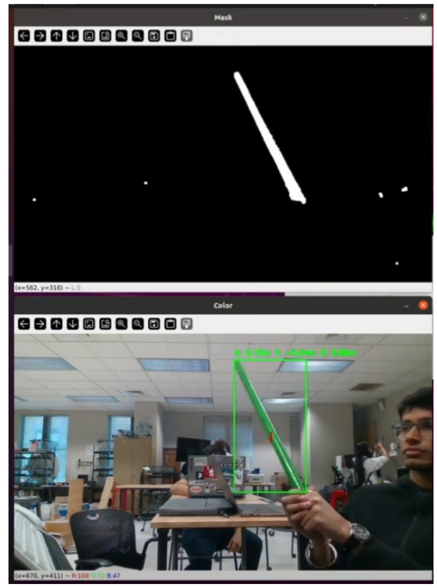

Fig. 4. Motion Tracking of Sword Centroid

visual field for tracking the human's motion over the duration of the engagement.

At present, our implementation enables the reliable detection and tracking of the fencing stick's centroid by utilizing HSV color thresholding techniques. This is facilitated by the distinctive green coloration of the PVC pipe, which we segment from the RGB video stream captured by the RealSense camera. Once the region of interest is isolated, the next stage in the perception pipeline involves extracting the corresponding depth information from the aligned depth map produced by the RealSense sensor suite. With both the 2D pixel location and the associated depth value available, we compute the 3D spatial coordinates of the fencing stick in the camera frame. These coordinates are then transformed into the robot's base coordinate frame using known extrinsic calibration parameters.

The resulting 3D pose estimate of the fencing stick is subsequently forwarded to the planning subsystem. This integration enables the robot to dynamically generate responsive trajectories in real time, allowing it to execute effective blocking or defensive maneuvers in response to the detected fencing motion of the human opponent.

C. Planning

The planning subsystem plays a critical role in enabling the robot to generate reactive motion trajectories that transition the robotic arm into a defensive posture in response to the detected movements of the human opponent. The planner not only facilitates this initial defensive behavior but is also designed to continuously mirror the human's motion in real time, thereby maintaining a dynamic and responsive interaction loop.

Upon receiving the estimated 3D pose of the fencing stick from the perception subsystem, the planner processes this input to compute an appropriate response strategy. This involves predicting the likely point of contact between the incoming stick and the robot defense area, modeled as a predefined virtual defense plane located in the workspace. To reduce the complexity of the entire planning problem and enable incremental development, we introduce a simplified preliminary subtask. In this sub-task, the robot focuses on mimicking the trajectory of the human-held stick in real time. This approach allows us to establish a reliable data processing pipeline and validate the system's ability to execute low-latency trajectory planning and execution.

The robot initiates active trajectory generation when the opponent's stick breaches a specified spatial threshold, located near the virtual defense plane. This spatial threshold functions as a trigger condition, signaling the need for the robot to transition from passive observation to active defense. Within this defensive motion logic, we implemented a mirroring strategy in which the robotic arm is programmed to replicate the human stick's motion precisely along the y and z axes, while inverting the motion along the x-axis (the x-axis is perpendicular to the image plane). This inversion is crucial for simulating a natural defensive response: When the human moves the stick toward the robot, the robot moves its arm toward the incoming trajectory, thus engaging the attack at an optimal angle.

The following shows the pseudo-code which describes the high-level logic of our pipeline:

Input: Real-time 3D pose of the opponent's baton in the camera frame
Initialize RealSense camera
while True **do**
    Capture RGB-D frame
    Estimate 3D centroid of the baton using HSV segmentation and depth map
    Publish estimated pose to ROS topic `/baton_pose_camera`
end while

```
Planner Node:
while True do
    Subscribe to /baton_pose_camera
    Transform pose to robot base frame using extrinsic calibration
    if x-position < threshold_x then        ▷ Check if opponent breaches defense plane
        break
    else
        Compute mirrored pose across the virtual camera plane
        Set fixed end-effector orientation
        Execute go_to_pose
    end if
end while
```

Additionally, to ensure that the orientation of the end effector aligns correctly for effective blocking, we integrated the quaternion-based rotational handling capabilities provided by the frankapy library. Using these built-in functions, the planner dynamically adjusts the rotation of the final robot joint, maintaining an orientation that ensures that the end effector remains in a perpendicular counterposition to the trajectory of the incoming stick. This configuration was key in achieving a functionally viable Minimum Viable Product (MVP), enabling robust and timely defensive actions within the constraints of our experimental framework.

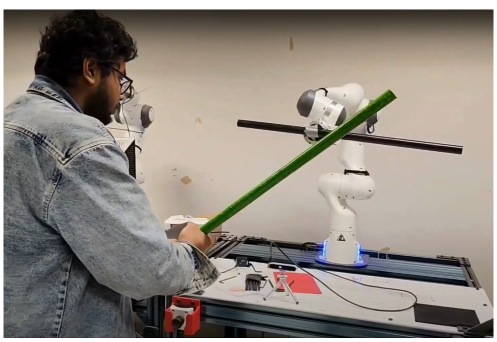

Fig. 5. Full System in Operation

IV. EVALUATION

To assess the effectiveness of our robotic fencing system, we performed a series of controlled experiments focused on evaluating four key performance metrics: perception accuracy, response latency, defensive maneuver success rate, and qualitative interaction feedback.

A. Perception Accuracy

We evaluated the stability of pose estimation for the human-held fencing baton during periods of minimal motion to understand how much the centroid values fluctuate under quasi-static conditions. Across 15 trials where the baton was held approximately still for 5 seconds, the estimated 3D centroid positions exhibited a standard deviation of approximately ± 1.1 cm in the x-y plane and ± 1.7 cm along the z-axis (depth), indicating mild jitter in the perceived location. This fluctuation is attributed primarily to depth quantization noise from the RealSense D415i sensor and occasional inconsistencies in HSV-based segmentation, which struggles under variable lighting or partial occlusion. While this level of jitter is tolerable for trajectory mirroring during continuous motion, it introduces uncertainty during precise defensive maneuvers that depend on consistent pose estimates. Future iterations could benefit from filtering techniques such as Kalman smoothing to reduce this noise and stabilize tracking during both static and dynamic conditions.

B. Response Latency

The latency between the detection of an incoming attack and the initiation of a robot response was a critical performance metric. We measured the end-to-end system latency from frame acquisition to trajectory execution trigger. The average latency across 15 trials was **165 ms**, with the major contributors being camera frame delay and planning pipeline overhead. This response time was found to be within acceptable limits for defensive maneuvers against medium-speed attacks, though it poses a challenge during rapid, high-velocity thrusts.

C. Defensive Maneuver Success Rate

We define a successful defensive maneuver as the robot's end-effector baton (the black sword) intersecting the predicted path of the opponent's sword before it reaches the hypothetical impact point at the robot's virtual defense plane. In the current MVP, the robot is not actively orienting its end-effector to optimally meet the opponent's sword at the point of contact—this capability has not yet been implemented or experimentally evaluated. As such, precise orientation-based interception is reserved for future work. Observed failure cases are primarily attributed to inaccurate pose estimation due to occlusions, or the robot exceeding virtual wall constraints designed to ensure safe operation.

D. Qualitative Assessment

In addition to quantitative metrics, we collected observational and subjective feedback from human participants interacting with the system. Participants noted that the robot's movements appeared fluid and responsive in most cases, particularly during imitation and mirroring phases. However, there were instances of noticeable lag and jitter during rapid angular transitions, especially when the baton changed orientation abruptly or other similar colored objects are present in the scene as the participants' baton.

E. Limitations

While the system performed reliably under moderate conditions, several limitations were identified:

- **Limited orientation control:** The robot's end-effector currently maintains a static orientation, limiting the ability to counter complex angled attacks.
- **Perception sensitivity:** HSV-based segmentation is sensitive to lighting changes and partial occlusions, occasionally leading to lost tracking or incorrect depth values.

- **Virtual wall constraints:** Safety constraints, while necessary, sometimes truncate valid defensive trajectories, especially during aggressive attacks near workspace limits.

F. Summary

The evaluation confirms that our system is capable of robust perception, timely reaction, and effective motion mirroring in a majority of fencing scenarios. With enhancements in end-effector orientation control and more robust filtering of pose estimates, the system holds promise as a platform for real-time, interactive human-robot fencing engagements.

V. CHALLENGES

During the development of our robotic fencing system, we encountered several challenges in perception, human-robot interaction (HRI), and system constraints.

A. Camera Interface and Perception Limitations

Our initial plan was to use the Microsoft Kinect for motion tracking because of its depth-sensing capabilities. However, due to unresolved compatibility issues with the lab PC, we had to pivot at the last minute to the Intel RealSense camera. Although RealSense performed reliably for basic depth-based tracking, it lacked the precision required to determine the exact point of contact between the opponent's sword and the robot's sword. This limitation made it difficult to verify successful defensive blocks in real time and restricted the granularity of feedback for physical interactions.

B. Human-Robot Interaction Instability

One of the major challenges in our system was dealing with the inherent instability in the human opponent's pose. Since the human holds the sword freely and engages in fast, unstructured motion, the detected pose of the sword often fluctuated, introducing noise in the trajectory prediction pipeline. This unpredictability affected the robot's ability to generate smooth and responsive trajectories. Although we considered filtering techniques such as Kalman or complementary filters, these were not implemented in the current iteration and remain a promising direction for future refinement.

C. Virtual Wall Constraints

To ensure safety, the robot operated within strictly defined virtual wall limits. While these constraints were critical in preventing accidental collisions or unsafe motion near the human, they also introduced limitations in trajectory planning. In several scenarios, the robot's intended motion was truncated or aborted due to exceeding these virtual bounds, especially during fast or aggressive attack simulations. This not only disrupted the overall interaction flow but also prevented the robot from executing high-speed trajectories necessary for realistic fencing engagements.

VI. STRETCH GOALS

As a natural extension of the motion mirroring capability developed in our minimum viable product, we are currently exploring the integration of *Iterative Learning Control (ILC)* to improve the performance and precision of motion tracking in repetitive sword movements. ILC is well-suited for systems that repeat tasks, as it uses tracking errors from previous iterations to refine the control input in subsequent trials. In our context—where sword swings often follow repeatable trajectories—ILC can help the robot end-effector adapt and improve its tracking fidelity over time, leading to smoother and more accurate mirroring.

Technically, ILC updates the control input for the $k+1$-th execution using the law:

$$u_{k+1}(t) = u_k(t) + L(e_k(t))$$

where $u_k(t)$ is the control input at iteration k, $e_k(t) = r(t) - y_k(t)$ is the tracking error, and L is the learning gain. For robotic manipulators, both joint-space and Cartesian-space ILC approaches have been shown to improve trajectory tracking performance, even under non-linear dynamics and time delays. Prior work has successfully applied ILC in human-robot interaction scenarios and imitation learning tasks [7], [8], [9].

At the time of writing, this feature is under active development and we do not yet have experimental outputs. However, preliminary setup and data collection are in progress. The ultimate goal is to allow the system to not only replicate motions more accurately with repetition but also to perform anticipatory behaviors and compensate for system latency or human variability.

VII. FUTURE WORK

While the current system effectively tracks the centroid of an attacking sword and mirrors its trajectory for reactive defense, several enhancements could significantly improve both the precision and realism of the interaction. We outline two key areas of future development:

A. Dynamic Orientation of the Robot's Sword

Currently, the robot mirrors the attacker's sword movement to position its own sword in the path of the incoming strike, maintaining a static end-effector orientation. However, for a truly effective defense in fencing—especially against angled or rotational attacks—the defending sword should not only be positioned at the correct point but also oriented perpendicularly to the attack vector. This requires implementing dynamic end-effector rotation, where the robot's wrist joint (last revolute joint) actively adjusts the robot's orientation.

To achieve this, we plan to estimate the full 3D direction vector of the attacker's baton based on sequential centroid positions and apply a rotational transform to the robot's end effector such that the baton held by the robot is orthogonal to the estimated path of attack. This adaptation would significantly increase the realism of the interaction and introduce more physically accurate blocking postures akin to real fencing techniques.

B. Trajectory Generation Triggered by Force Feedback

The robot currently initiates a defensive trajectory once the opponent's sword crosses a spatial threshold near a virtual defense plane. While effective in a controlled setting, this approach lacks adaptability to variations in attack force and tempo. A more robust strategy would involve equipping the robot with force or torque sensors—either in the joints or at the end-effector—to detect the onset of contact or pre-contact pressure indicative of an impending strike.

Such a system would allow the robot to initiate trajectory planning not solely based on visual thresholds but in combination with tactile cues, increasing responsiveness and safety. This hybrid approach of using both vision and force feedback can enable the robot to detect and respond to attacks earlier and more reliably, paving the way for real-time parries or deflections that depend on human force input rather than just position.

VIII. CONCLUSION

In this work, we presented Project Yoriichi, a robotic fencing system that combines real-time visual tracking with responsive trajectory generation to enable interactive human-robot engagements. This project demonstrates the feasibility of implementing low-cost, vision-based robotic systems for dynamic, high-speed tasks like fencing, contributing to the broader fields of robotic sports training and human-robot collaboration. While the current implementation focuses on centroid-based tracking and positional mirroring, future enhancements—including dynamic sword orientation and force-triggered responses—will further improve the fidelity and responsiveness of the system. These advancements aim to create a more intuitive, physically grounded interaction model, moving closer to true robotic sparring capabilities.

REFERENCES

[1] M. Filip, "Advanced Motion Technologies for Supporting Fencing Athletes," *Journal of Sports Medicine & Doping Studies*, vol. 14, no. 03, 2024.

[2] K. Zhang, M. Sharma, J. Liang, and O. Kroemer, "A Modular Robotic Arm Control Stack for Research: Franka-Interface and FrankaPy," Robotics Institute, Carnegie Mellon University, 2020.

[3] Vicon, "Robot Tracking with Motion Capture for Robotics," Vicon, 2025.

[4] A. Namiki and F. Takahashi, "Motion Generation for a Sword-Fighting Robot Based on Quick Detection of Opposite Player's Initial Motions," *Journal of Robotics and Mechatronics*, vol. 27, pp. 543–551, 2015.

[5] A. Ghadirzadeh, A. Maki, D. Kragic, and M. Björkman, "Deep Predictive Policy Training using Reinforcement Learning," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[6] R. Rahmatizadeh, P. Abolghasemi, L. Bölöni, and S. Levine, "Vision-Based Multi-Task Manipulation for Inexpensive Robots Using End-to-End Learning from Demonstration," *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[7] R. Freeman and T. Parisini, "Iterative Learning Control: A Brief Survey and New Results," *Annual Reviews in Control*, vol. 35, no. 2, pp. 232–243, 2011.

[8] S. Krishnan and P. Bérard, "Learning Repetitive Motions with Iterative Learning Control for Human-Robot Collaboration," *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[9] H.-S. Ahn, Y. Chen, and K. L. Moore, "Iterative Learning Control: Brief Survey and Categorization," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1099–1121, 2007.