Automated Tennis Swing Analysis Using Pose Estimation and Optical Flow

Jeffrey Liu

Abstract—This project aims to help tennis players improve their swing mechanics by comparing their wrist trajectories and speeds to those of professional players. Using pose estimation via MediaPipe and optical flow for motion tracking, I analyze video recordings of myself and a professional player with similar physical attributes and style. The trajectories and speeds are visualized and compared, with quantitative metrics such as dynamic time warping, trajectory angles, and speed profiles calculated to assess performance. The results demonstrate the feasibility of using computational imaging techniques for deep analytical swing analysis, providing quantifiable and visual feedback for players that cannot be detected with the naked eye.

Index Terms—Pose Estimation, Tennis Swing Analysis, MediaPipe, Optical Flow, Computer Vision

1 Introduction

Tennis players often struggle to improve their swing mechanics due to the lack of accessible and affordable analysis tools. Professional coaching and motion capture systems, while effective, are expensive and inaccessible to most players. This project addresses this gap by leveraging computer vision and computational imaging techniques to analyze and compare swing trajectories and speeds. My method takes two videos—one of the user and one of a professional or reference player—and uses MediaPipe for pose estimation and optical flow for motion tracking to provide an efficient and simple alternative to traditional methods. The goal is to help players identify discrepancies in their swing mechanics that may be difficult to detect without visual aids, enabling them to improve their performance through actionable feedback.

2 Related Work

Traditional swing analysis relies on subjective coaching or expensive motion capture systems. While these methods are effective, they are not accessible to most players. Recent advances in computer vision and computational imaging have introduced low-cost, automated alternatives for tennis swing analysis. These include 2D and 3D pose estimation, optical flow-based motion tracking, and real-time graphics simulations.

MediaPipe, a widely used framework for 2D pose estimation, provides real-time, accurate detection of body landmarks without the need for specialized hardware. It has been applied to various sports analysis tasks, including tennis, to track player movements and analyze swing mechanics. However, 2D pose estimation is limited by its inability to capture depth information, which is crucial for a complete understanding of swing dynamics [2].

• *Jeffrey Liu is with the Department of Electrical Engineering, Stanford University, Stanford, United States.*
E-mail: liujy3@stanford.edu

To address the limitations of 2D methods, 3D pose estimation techniques have been developed. These methods use multiple cameras or depth sensors to reconstruct the player's motion in three dimensions. For example, systems like OpenPose and Vicon provide detailed 3D models of player movements, enabling more accurate analysis of swing mechanics. However, these systems often require expensive equipment and complex setups, making them inaccessible to most players [5].

Optical flow techniques, such as the Farnebäck method, have been widely used for motion tracking in video analysis. These methods estimate the motion of objects between consecutive frames, making them suitable for tracking the racket and player movements. Optical flow is computationally efficient and works well with single-camera setups, but it can struggle with occlusions and fast motions [1].

Real-time graphics and simulations have been used to create virtual environments for tennis training and analysis. For example, Unity and Unreal Engine have been employed to develop interactive tennis simulations that provide real-time feedback on player performance. These systems often integrate motion capture data to create realistic animations and analyze swing mechanics. While powerful, they require significant computational resources and are primarily used in professional settings [6].

Another approach to tennis swing analysis involves the use of wearable sensors, such as accelerometers and gyroscopes, attached to the player's body or racket. These sensors provide precise measurements of motion and force, enabling detailed analysis of swing mechanics. However, they are often intrusive and expensive, limiting their accessibility [3].

My work builds on these advances by combining 2D pose estimation (MediaPipe) and optical flow (Farnebäck method) to provide an automated and affordable solution for tennis swing analysis. Unlike 3D pose estimation and real-time simulations, this method requires only a single camera and does not rely on specialized hardware, making it accessible to a wider audience. By focusing on wrist trajectory and speed analysis, we provide actionable feedback to players without the need for expensive equipment or complex setups.

3 THEORY/METHOD

My method involves four main steps: (1) pose estimation using MediaPipe to detect key body points, (2) tracking the left wrist trajectory across video frames, (3) calculating wrist speed using optical flow, and (4) comparing the trajectories and speeds of an amateur and a professional player. The trajectories and speeds are visualized using OpenCV, and quantitative metrics such as average error, maximum deviation, and speed profiles are calculated to assess performance.

3.1 Pose Estimation

I use MediaPipe's pose estimation model to detect key body points in each frame of the video as shown in Figure 1,. The model provides specific joints as landmarks, including the left wrist, which is the focus of the analysis.

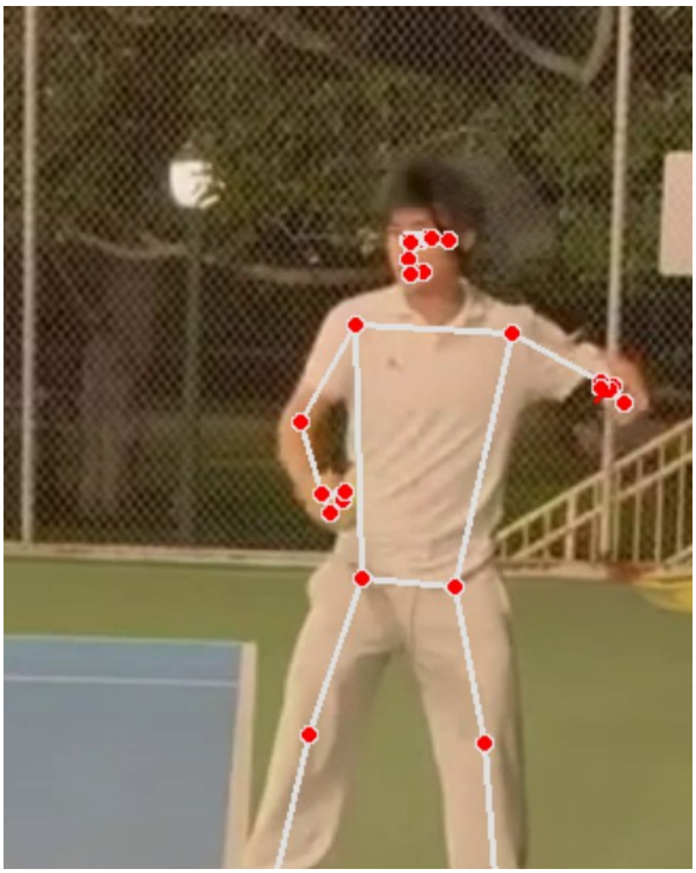

Fig. 1. Use of Mediapose to mark key joints for pose estimation.

3.2 Wrist Trajectory Tracking

The left wrist's position is tracked across all frames of the video. The trajectory is represented as a sequence of (x, y) coordinates in pixel space. To handle missing detections, I use the last known position of the wrist. To track the left wrist across video frames, I represent its position as

$$w(t) = (x(t), y(t))$$

in pixel space. If the wrist is detected in the tth frame, I set

$$w(t) = \hat{w}(t);$$

otherwise, I use the last known position:

$$w(t) = \begin{cases} \hat{w}(t), & \text{if detected in frame } t, \\ w(t-1), & \text{otherwise.} \end{cases}$$

This method ensures a continuous trajectory even in the presence of intermittent detection failures.

3.3 Speed Calculation

The speed of the wrist is calculated using optical flow. The displacement of the wrist between consecutive frames is measured, and the speed is computed as the magnitude of the displacement vector, scaled by the frame rate f to provide real-world units (pixels per second). Let $\mathbf{p}(t) = (x(t), y(t))$ denote the position of the left wrist in frame t. The displacement vector between consecutive frames is defined as:

$$\Delta\mathbf{p}(t) = \mathbf{p}(t) - \mathbf{p}(t-1) = \big(x(t) - x(t-1),\ y(t) - y(t-1)\big).$$

The magnitude of this displacement, representing the movement in pixels per frame, is given by:

$$\|\Delta\mathbf{p}(t)\| = \sqrt{\big(x(t) - x(t-1)\big)^2 + \big(y(t) - y(t-1)\big)^2}.$$

To obtain the speed $v(t)$ in pixels per second, I scale the displacement by the frame rate f (in frames per second):

$$v(t) = f \cdot \|\Delta\mathbf{p}(t)\| = f \cdot \sqrt{\big(x(t) - x(t-1)\big)^2 + \big(y(t) - y(t-1)\big)^2}.$$

3.4 Trajectory and Speed Comparison

The trajectories and speeds of the amateur and professional players are compared by calculating the Euclidean distance between corresponding points and analyzing the speed profiles. The trajectory angle, dynamic time warping, and speed differences are computed to quantify the discrepancies. I compare the trajectories and speeds of the amateur and professional players using the following procedures.

Trajectory Analysis: Let $\mathbf{p}(t) = (x(t), y(t))$ denote the position of the left wrist in frame t. The overall displacement (net distance) for a swing is computed as:

$$d = \|\mathbf{p}(T) - \mathbf{p}(1)\| = \sqrt{\big(x(T) - x(1)\big)^2 + \big(y(T) - y(1)\big)^2},$$

and the swing angle is given by:

$$\theta = \arctan\left(\frac{y(T) - y(1)}{x(T) - x(1)}\right).$$

Here, T represents the final frame of the swing.

Trajectory Comparison: For the amateur (user) and professional (pro) players, let $d_{\text{user}}, \theta_{\text{user}}$ and $d_{\text{pro}}, \theta_{\text{pro}}$ denote the net displacement and swing angle, respectively. The discrepancies are quantified as:

$$\Delta d = |d_{\text{user}} - d_{\text{pro}}|, \quad \Delta\theta = |\theta_{\text{user}} - \theta_{\text{pro}}|.$$

Speed Profile Analysis: The instantaneous speed is computed by first measuring the displacement between consecutive frames:

$$\|\Delta\mathbf{p}(t)\| = \sqrt{\big(x(t) - x(t-1)\big)^2 + \big(y(t) - y(t-1)\big)^2},$$

and then scaling by the frame rate f (in frames per second) to obtain the speed in pixels per second:

$$v(t) = f \cdot \|\Delta \mathbf{p}(t)\|.$$

The average speed over N valid frames is then given by:

$$\bar{v} = \frac{1}{N} \sum_{t=1}^{N} v(t).$$

In my implementation, these computations are performed in the following code sections:

- The functions `analyze_swing` and `compare_swings` (in the final analysis block of `main()`) compute and compare the net displacement and swing angle.
- The instantaneous speeds are computed in `process_video` using `measure_optical_flow_speed`, scaled by the frame rate, and subsequently smoothed with `smooth_speed_curve`. The speed profiles are then visualized by plotting the smoothed speeds.

4 Analysis & Evaluation of Different Methods

4.1 Comparison with Existing Methods

To evaluate the effectiveness of this method, I compare it to several existing approaches for tennis swing analysis, including wearable sensors, 3D pose estimation, and manual analysis. Below, I discuss the strengths and limitations of each method and highlight how this new approach addresses their shortcomings.

4.1.1 Wearable Sensors

Wearable sensors, such as accelerometers and gyroscopes, provide precise measurements of motion and force, enabling detailed analysis of swing mechanics. However, these sensors are often intrusive, expensive, and require specialized equipment, making them inaccessible to most players. My method eliminates the need for wearable sensors by using computer vision techniques, making it non-intrusive and affordable.

4.1.2 3D Pose Estimation

3D pose estimation techniques, such as OpenPose and Vicon, provide detailed 3D models of player movements, enabling accurate analysis of swing mechanics. However, these systems require multiple cameras or depth sensors, making them expensive and complex to set up. My method uses a single camera and does not require specialized hardware, making it more accessible and cost-effective.

4.1.3 Manual Analysis

Manual analysis by coaches provides subjective insights into swing mechanics and is widely used in professional settings. However, this approach is time-consuming, subjective, and relies heavily on the coach's expertise. My approach automates the analysis process, providing objective and quantitative feedback without the need for manual intervention.

4.1.4 Real-Time Graphics and Simulations

Real-time graphics and simulations, such as those developed using Unity or Unreal Engine, provide interactive environments for training and analysis. However, these systems require significant computational resources and are primarily used in professional settings. My method is computationally efficient and can run on standard hardware, making it suitable for casual players and coaches.

5 Experimental Results and Analysis

In order to validate that this approach is working correctly, I conducted an experiment to evaluate the performance of the pose estimation and swing analysis framework. Two video recordings were used for this purpose: one of an amateur player ("liu_forehand.mp4") and one of a professional player ("iwa_forehand.mp4"). The experiments were performed using a custom Python script that integrates MediaPipe for pose estimation, optical flow for speed computation, and various post-processing steps to analyze the swing trajectories.

First, MediaPipe library is integrated with each video to have reliable marking. I want to track the left wrist as the mark for drawing out the swing path because both the user's and professionals' swing is left-handed. One observation when first implementing is that when drawing out the swing path, there can be slight missing detections throughout the swing. This issue has been addressed using the method described in Section 3.2. The swing paths for both the user's swing and the professional player's swing are mapped onto normalized coordinates to highlight similarities and differences.

One major challenge is mapping it so that the initial position of each swing has the same reference. This is resolved by normalizing the swing trajectories to account for different body sizes by using the shoulder-to-shoulder distance as a reference scale. After normalization, the starting points of the user's and pro's swing trajectories are aligned. This ensures that the comparison begins from the same initial position. Finally, the normalization and alignment steps are combined to ensure that the swing trajectories are both scale-invariant and aligned at the starting point.

Another challenge is that the video provided may not start at the same frame. As shown in figures 4 and 5 of the frame by frame comparison, there is some trial and error testing for the video of my swing to start around frame 20 to match with the same swing position as the professional. This is because in the user video, the initiation of the swing starts later than the professional's swing video.

The next step is measuring the swing speed throughout the entire swing. The calculation to do this is described in section 3.3. One challenge is that there are sudden glitches close to 0 pixels/s for certain parts of the swing that should not be the case. The computed speed for each frame is later smoothed using a Savitzky–Golay filter in the function `smooth_speed_curve` to help solve this problem. Furthermore, I used a filter to preserve the high speeds but filtered out the low speeds where the swing is actually accelerating. This made the graph of the swing speed even smoother as shown in Figure 4, I evaluated this method by comparing the wrist trajectories and speeds of my swing to those of a professional player named Sato Iwabuchi, who has a similar physique to make the comparison more realistic. In terms of the swing speed graph comparison in Figure 2, the change of speeds throughout the swing path makes sense. In the beginning of the swing, the left hand starts moving from the ready position as shown in the slow speed up in the early frames. Then as each player is tracking the ball to find the best position to strike, there is a slight pause showing the dip in speed around frame index 30. Then both players will accelerate and swing at the ball with the maximum speed, both being at the contact of the tennis ball in around frame index 35. Then towards frame index 40, the swing has completed as the left hand and arm wraps around to the right shoulder. This also matches with the pictures showing proof that this graph makes sense.

Fig. 2. Frame by frame of my swing with recorded speed in pixels/s.

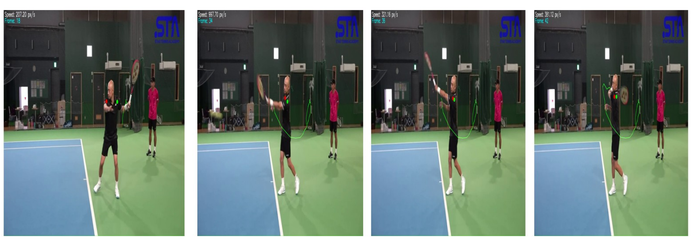

Fig. 3. Frame by frame of a professional player's (Sato Iwabuchi) swing with recorded speed in pixels/s.

TABLE 1
Summary of Key Swing Metrics for User vs. Professional

Metric	User	Pro
Mean Velocity (px/s)	193.70	363.31
Mean Acceleration (px/s^2)	57.91	88.18
Net Distance & Angle		
Distance (px)	3.08	2.60
Angle (degrees)	187.07	167.92
Difference (User - Pro)		
Distance (px)	0.48	
Angle (degrees)	19.15	

As shown in Figure 3, the swing trajectories are pretty similar, with some differences. The swing of the amateur (Jeffrey Liu) appears to be bigger and less compact compared to the professional swing. To further validate this, the net distance is calculated in pixels showing that the user's swing is 3.08 pixels compared to the pro's swing of 2.60 pixels. This is not a significant issue but in tennis, having a more compact swing usually means fewer errors and better preparation, especially at higher levels where tennis balls are coming much faster than recreational tennis. In terms of the angle of the swing path, both swings are pretty similar with the user swing slightly more vertical as shown in Table 1 of the net angle with the user's swing angle to be 187.07 degrees and the pro's swing is 167.92 degrees.

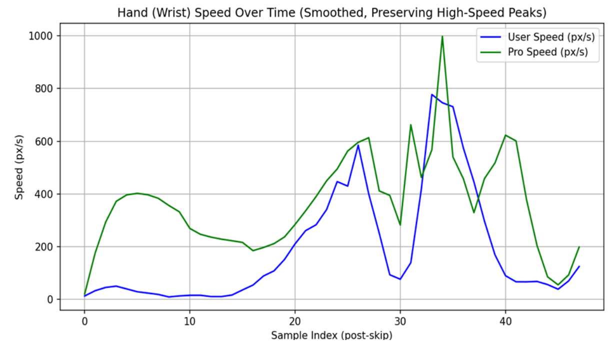

Fig. 4. Overlay of amateur (blue) and professional (green) swing speeds.

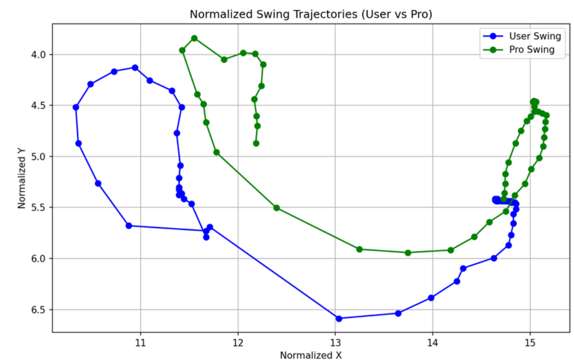

Fig. 5. Overlay of amateur (blue) and professional (green) wrist trajectories.

Additional data are calculated and recorded as mentioned earlier in section 3. The mean velocity and mean acceleration of the pro's swing are significantly faster than mine, with the pro's mean velocity and acceleration to be 193.70 pixels per second and 57.91 pixels per second squared respectively. While the user's swing speed is only 193.70 pixels per second and 57.91 pixels per second squared. This can help recreational players like me realize that not only does the form have to be good, but also how you are accelerating and impacting the ball matters as that determines how fast you can hit as well.

5.1 Experimental Setup

The videos were processed with the following settings:

- **Input Videos:** Amateur video (`liu_forehand.mp4`) and professional video (`iwa_forehand.mp4`).
- **Frame Resolution:** All frames were resized to 640×360 pixels for consistency.
- **Capture Interval:** Frames were sampled every 0.05 seconds.
- **Pose Estimation:** MediaPipe's pose model was used to detect key landmarks, and the left wrist was extracted using the `track_racket` function.

6 Discussion, Limitations, and Future Work

This approach successfully tracks and compares wrist trajectories and speeds, but it is sensitive to occlusions and requires decent video input quality in order for the tracking to be accurate. Another limitation is that in order to have an accurate comparison, the same stroke and camera angle as well as similar starting time should be taken into account for more accurate results. Also, the dataset that I test for this experiment is very small. Thus, for future work, I should try with more videos and more repeated strokes for better analysis. In addition, I could include tracking additional joints, such as the elbow and shoulder angles, to provide a more comprehensive analysis. Real-time feedback could also be implemented to assist players during practice. Additionally, integrating machine learning models to predict swing efficiency based on the tracked data could further enhance the analysis. Another aspect to look into is 3D-pose estimation as it can provide additional feedback like the racket orientation upon contact and torso rotation that are helpful in assessing tennis stroke and speed.

7 Conclusion

In this project, I demonstrated the feasibility of using computational imaging techniques, particularly optical flow and pose estimation for affordable and accessible tennis swing analysis. My method provides actionable feedback for players and coaches, bridging the gap between amateur and professional training. Future work will focus on improving robustness and expanding the analysis to include additional joints and real-time feedback.

Acknowledgments

I would like to thank Qingqing Zhao for mentoring this project.

References

[1] Brox, T., et al. (2004). High accuracy optical flow estimation based on a theory for warping. In European Conference on Computer Vision (ECCV).

[2] Cao, Z., et al. (2019). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] Kelly, P., O'Connor, N. E. (2012). Visualisation of tennis swings for coaching. In 2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (pp. 1-4). Dublin, Ireland.

[4] Wang, C. (2023). Method for capturing tennis training actions under machine vision images. In 2023 International Conference on Computer Science and Automation Technology (CSAT) (pp. 409-413). Shanghai, China.

[5] Xu, T., Li, Z., Yuan, M., Zheng, Z., Zhang, J., Kuai, X. (2023). Three-dimensional spatiotemporal reconstruction and feature analysis of table tennis movement enhanced by multi-view computer vision. In 2023 3rd International Conference on Information Technology and Contemporary Sports (TCS) (pp. 60-68). Guangzhou, China.

[6] Zhang, H., Yuan, Y., Makoviychuk, V., Guo, Y., Fidler, S., Peng, X. B., Fatahalian, K. (2023, August). Learning physically simulated tennis players from broadcast videos. ACM Transactions on Graphics (SIGGRAPH).