

KaratePose-AI

Kinematic Refinement of Coordinate-Based Human Pose Estimation via a Physics-Informed Filter Engine for Elite Martial Arts Analytics

Undergraduate Research & Development Proposal — Revision 2 (Literature-Grounded)

A Two-Phase Study on Rigid-Garment Occlusion and Event-Triggered Kinematic Correction in Martial Arts Pose Estimation

Degree Programme: BSc (Hons) Artificial Intelligence — University of Plymouth / NSBM

Candidate Venues: To be determined based on scope and peer-review fit

1. Executive Summary & Objective

State-of-the-art (SOTA) human pose estimation (HPE) models — Google MediaPipe and YOLOv8-Pose — deliver strong accuracy on standard benchmark datasets featuring form-fitting apparel. Prior work has separately studied martial arts pose estimation (in standard apparel), loose-clothing occlusion (in winter sports), and Kalman-filtered biomechanical tracking (in sprinting) — but not the specific combination this project targets: rigid, heavyweight canvas fabric (e.g. a 14oz Karate Gi) folding under high-velocity kinetic transitions. This is a mechanically distinct occlusion pattern from loose drape-type fabric, and it has not been benchmarked or corrected for in the literature reviewed to date.

This project addresses that gap in two phases:

- 1. Quantify the degradation of standard coordinate-prediction models (MediaPipe, YOLOv8-Pose) under rigid-fabric occlusion and high-velocity transitions (Paper 1).
- 2. Develop and validate an event-triggered, physics-informed correction filter — an Extended Kalman Filter (EKF) that activates on confidence collapse rather than smoothing continuously — to reconstruct true joint trajectories and compute Center of Gravity (CoG) (Paper 2).

2. Introduction & Problem Statement

SOTA computer-vision HPE frameworks suffer measurable performance degradation when exposed to two concurrent real-world variables: (1) high-velocity kinetic transitions, which introduce localized motion blur, and (2) complex material occlusion, where rigid canvas fabric folds and alters body silhouettes, collapses visual joint markers, and introduces false edge boundaries. This differs mechanically from the loose-fabric draping studied in winter-sports HPE literature: rigid canvas creates hard, false edges rather than soft occlusion. This research formulates a physics-informed post-processing pipeline that stabilizes tracking and produces reliable athletic metrics under these conditions — using consumer-grade hardware only.

3. Related Work & Positioning

A review of adjacent literature shows this project sits at an intersection that has not been directly addressed, though each neighboring piece is well established individually:

Prior Work	What It Covers	What It Doesn't Cover
Echeverria & Santos (2021), Sensors	Pose estimation (OpenPose) applied to karate kumite to model psychomotor performance.	No garment variable; assumes standard training apparel.
Zhang et al. (2017) — MADS dataset	Multi-view + stereo-depth benchmark including Karate and Tai-chi performed by martial arts masters.	Targets general self-occlusion and fast articulation, not garment-specific occlusion.
Dael et al. (2026), ICAAIC	Direct MediaPipe vs. YOLOv8 comparison on scenarios including martial techniques.	Measures FPS / CPU / memory — not accuracy degradation from clothing.
Handball Kalman-smoothing study (2022), J. Imaging	Per-axis Kalman smoothing applied to noisy pose-estimation keypoints in sport.	Continuous smoothing for general noise, not an event-triggered response to occlusion-driven confidence collapse.
Sprinting CoM study (2021), Sensors	The closest direct ancestor: Kalman-smoothed pose keypoints → de	Applied to sprinting biomechanics, not to garment occlusion or martial

Prior Work	What It Covers	What It Doesn't Cover
	Leva segment model → center-of-mass, validated against marker-based ground truth.	arts.
Nakai et al. (2022), loose-clothing ground truth	HFR camera + blinking LEDs under loose clothing to obtain verified ground truth for skiing/snowboarding apparel.	Addresses loose drape-type occlusion, not rigid canvas folding; not applied to combat sports.

Positioning statement: this project does not claim to invent martial-arts pose estimation, Kalman-filtered biomechanical tracking, or clothing-occlusion research — each already exists. The contribution is narrower and more defensible: characterizing the specific compounding failure of rigid-canvas occlusion under high-velocity kinetics, and correcting it with an event-triggered (confidence-gated) filter rather than the continuous smoothing used in prior sports-HPE Kalman applications.

4. Research Objectives

#	Objective
1	Establish a controlled dual-garment dataset isolating rigid canvas-fabric occlusion from high-velocity martial arts kinetics, extending existing karate pose-estimation work (which assumes standard apparel) and existing loose-clothing occlusion work (which addresses draping, not folding).
2	Quantify degradation using confidence-score volatility and spatial jitter, with a pre-registered statistical comparison (Set A vs. Set B).
3	Construct a biomechanical framework (de Leva segment model) to compute live Center of Gravity and Stance Depth, following the validated approach used in markerless sprinting biomechanics.
4	Design and validate an event-triggered Extended Kalman Filter that rejects low-confidence coordinates and projects joint location via momentum — distinct from the continuous smoothing used in prior sports-HPE Kalman applications — benchmarked against a linear-interpolation baseline.
5	Publish two manuscripts — a benchmarking study and an algorithmic optimization methodology — with an honest discussion of dataset scale as a limitation.

5. Methodology & System Architecture

5.1 Data Capture Protocol

A controlled dataset will be captured with an iPhone 15 Pro Max on a fixed tripod, using two settings:

- 4K at 60 FPS — standard tracking and inference testing.
- 1080p at 240 FPS — slow-motion baseline for visual cross-checking during blur.

Two matched conditions will be filmed performing foundational Shotokan Katas (Oi-zuki, Gedan-barai, Zenkutsu-dachi transitions):

- Set A — Control: tight, lightweight athletic apparel.
- Set B — Experimental: heavyweight 14oz canvas Gi — the occlusion condition.

Dataset scale: where feasible, footage will be collected across multiple sessions and, ideally, more than one practitioner. A single-subject, single-session dataset is treated explicitly as a pilot/feasibility scope in Section 8 (Limitations), not disguised as a generalizable result.

5.2 Strengthening Ground Truth

Tight clothing (Set A) is a proxy for ground truth, not verified ground truth — MediaPipe/YOLO can still err on unoccluded footage. Following the approach used in loose-clothing HPE ground-truth work, a small number of high-contrast markers (e.g. colored tape or small reflective tags) will be placed at key joints for a subsample of takes in both conditions, enabling manual frame-by-frame verification independent of any pose-estimation model. This gives Paper 1 a semi-independent accuracy check beyond “Set A looks smooth.”

5.3 Coordinate Ingestion Pipeline

A Python pipeline (OpenCV, MediaPipe, Ultralytics YOLOv8) streams both video sets frame-by-frame and logs structural joint arrays into structured DataFrames:

- MediaPipe: 33 landmarks per frame, each with (x, y, z) coordinates and a visibility/confidence score (v).
- YOLOv8-Pose: 17 keypoints per frame with bounding-box coordinates and a detection confidence score (c).
- All frames are indexed against key joints: Ankle, Knee, Hip, Shoulder, Elbow, Wrist.

5.4 Phase 1 — Material Occlusion Benchmark (Paper 1)

Two degradation vectors are isolated and quantified:

- Confidence Score Volatility — mapping drops in detection confidence (v or c) for the hip and knee joints during deep-stance transitions.
- Spatial Jitter — the frame-to-frame Euclidean displacement of each tracked coordinate:

$$Jitter = \text{sqrt}[(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2]$$

5.5 Statistical Analysis Plan

Defined before data collection, not after:

- Primary comparison: jitter magnitude, Set A vs. Set B, per joint — Wilcoxon signed-rank test (paired, non-parametric; jitter distributions are expected to be right-skewed).
- Effect size reported alongside significance (rank-biserial correlation or Cohen's d on log-transformed jitter).
- Confidence-score volatility compared using the same paired non-parametric approach.
- Phase 2 evaluation: EKF-corrected trajectories vs. (a) raw Set B output and (b) a linear-interpolation baseline — not EKF vs. nothing.

5.6 Phase 2 — Physics-Informed Kinematic Refinement (Paper 2)

Biomechanical CoG Mapping: the human body is modeled as a linked-segment system using de Leva's (1996) anatomical mass percentages (e.g. Torso ≈ 43.46%, Thigh ≈ 14.16%, Lower Leg ≈ 4.33%), following the same approach validated for markerless center-of-mass tracking in sprinting biomechanics. Global Center of Gravity for a given frame is:

$$X_CoG = \Sigma(x_i \cdot w_i) \quad Y_CoG = \Sigma(y_i \cdot w_i)$$

where x_i , y_i is the midpoint of body segment i and w_i is its relative anatomical mass weight.

Event-Triggered State Correction: unlike prior sports-HPE applications that apply Kalman smoothing continuously across every frame, this filter is designed to activate specifically on occlusion events. While confidence remains high, the filter

updates its velocity model from the raw AI output. When a joint's confidence score drops below an empirical threshold ($T < 0.50$), the filter rejects the raw coordinate and instead projects the joint's location forward using its established momentum vector. Threshold T will be tuned on a held-out portion of the dataset, not fixed a priori. This is evaluated against a linear-interpolation baseline to demonstrate the momentum-based approach is doing more than simple gap-filling.

6. Hardware & Software Infrastructure

The entire development loop runs on existing consumer hardware, requiring no additional budget or university compute infrastructure:

- Data Capture: iPhone 15 Pro Max (60/240 FPS).
- Local ML Training & Inference: ASUS ROG G513 (AMD Ryzen 7 6800H, NVIDIA RTX 3060 6GB VRAM) — CUDA-accelerated YOLOv8 pipelines.
- Analysis & Composition: MacBook Air M4 — Jupyter Notebooks for statistical analysis, LaTeX manuscript drafting.
- Software Stack: Python — OpenCV, MediaPipe, Ultralytics YOLOv8, Pandas, NumPy, SciPy.

Deployability note: lightweight consumer-grade models (MediaPipe, YOLOv8) are used deliberately, not because heavier transformer-based pose models are unknown to the field — they are, and tend to outperform on accuracy in adjacent sports benchmarks. The choice reflects a real dojo-deployment use case (no GPU cluster, no motion-capture lab) and is stated as a design decision in the manuscript, not left implicit.

7. Expected Academic Contributions

- Manuscript 1 (System Evaluation): a benchmark quantifying pose-model degradation under rigid-fabric occlusion combined with high-velocity martial arts kinetics — a compounding failure mode not isolated in prior karate-HPE or clothing-occlusion literature.
- Manuscript 2 (Algorithmic Solution): an event-triggered EKF correction method, benchmarked against a linear-interpolation baseline, extending the Kalman+de-Leva pipeline validated in sprinting biomechanics to a garment-occlusion setting.

Venue placement will depend on peer-review outcomes and is not fixed at proposal stage.

8. Limitations & Threats to Validity

- Dataset scale: a single-subject dataset (if that is the final scope) limits generalizability; this will be stated explicitly as a pilot/feasibility study, not implied to be population-level.
- Ground truth: Set A is a proxy, not verified ground truth; the marker-assisted subsample (Section 5.2) partially, not fully, mitigates this.
- Model scope: only MediaPipe and YOLOv8-Pose are benchmarked; transformer-based models (e.g. ViTPose) are known to outperform both in adjacent sports domains and are noted as future work rather than omitted silently.
- Threshold tuning: the $T < 0.50$ confidence threshold will be empirically tuned rather than assumed correct from the outset.

9. Project Timeline

Phase	Months	Goal	Key Deliverable
1 — Foundations & Environment Setup	1–3	Master Python data handling; stand up the video/coordinate pipeline	Working extraction script; annotated

Phase	Months	Goal	Key Deliverable
		(pandas, OpenCV, MediaPipe); finish literature review.	bibliography
2 — Video Acquisition & Baseline Dataset	4–6	Capture the controlled dual-garment dataset across multiple sessions (and additional practitioners if available); apply joint-marker ground-truth aid.	Complete Set A / Set B video corpus + marker-verified subsample
3 — Paper 1: Benchmarking Analysis & Write-up	7–12	Run extraction on both sets; run pre-registered statistical tests; draft manuscript.	Manuscript 1 drafted
4 — Biomechanical Integration	13–16	Implement Center-of-Gravity and Stance Depth tracking on the clean Set A baseline; build linear-interpolation baseline filter.	Working CoG overlay + baseline filter
5 — Physics-Informed Filter & Paper 2	17–22	Build and validate the Extended Kalman Filter on occluded Set B data; benchmark against the baseline; draft manuscript.	Manuscript 2 drafted
6 — Review, Revision & Submission	23–24	Incorporate supervisor/peer feedback; finalize both manuscripts for submission.	Both manuscripts submitted

10. References

- de Leva, P. (1996). Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. *Journal of Biomechanics*, 29(9), 1223–1230.
- Echeverria, J., & Santos, O. C. (2021). Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation. *Sensors*, 21(24), 8378.
- Zhang, W., Liu, Z., Zhou, L., Leung, H., & Chan, A. B. (2017). Martial Arts, Dancing and Sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation. *Image and Vision Computing*, 61, 22–39.
- Dael, F. A., et al. (2026). Comparative Evaluation of MediaPipe and YOLOv8 for Real-Time Pose Estimation. *Proceedings of ICAAIC / Springer*.
- Journal of Imaging (2022). 3D Pose Estimation and Tracking in Handball Actions Using a Monocular Camera. *Journal of Imaging*, 8(11), 308.
- Sensors (2021). Can Markerless Pose Estimation Algorithms Estimate 3D Mass Centre Positions and Velocities during Linear Sprinting Activities? *Sensors*, 21(8), 2889.
- arXiv:2212.04820 (2022). Pose Estimation for Human Wearing Loose-Fitting Clothes: Obtaining Ground Truth Posture Using HFR Camera and Blinking LEDs.
- Lu, F. AiShifu: AI Karate Pose Trainer Using Human Pose Estimation. *Journal of Student Research*.
- Note: this list reflects a preliminary literature scan to establish positioning and should be expanded into a full, systematically-searched related-work section (with complete author lists and verified venues) before manuscript submission.*

11. Appendix — Self-Guided Study Index

Suggested search terms to accelerate independent, ahead-of-coursework learning:

Python Engineering

- "Python pandas DataFrame tutorial"
- "OpenCV read video frame by frame"
- "How to format numpy arrays"

Computer Vision

- "YOLOv8 pose estimation inference python"
- "MediaPipe holistic landmark mapping"
- "Object detection confidence score extraction"

Mathematics, Physics & Statistics

- "De Leva human body segment parameters"
- "Kinematic velocity vectors from tracking coordinates"
- "1D Kalman filter python implementation from scratch"
- "Wilcoxon signed-rank test paired samples python"