Robotics & Intelligent Systems Expo (RISEx 2025), Nov. 13-14, 2025, Edmonton, Canada

A LOW-COST SYSTEM FOR MARKERLESS BODY PART INTERACTION DETECTION IN PHYSICAL THERAPY

Yashar Talebirad[1,2], Simon Nakatsu[1,2], and Brenda Luo[3]
[1] Faculty of Science, University of Alberta, Edmonton, Canada.
[2] Lunate Labs, Calgary, Canada
Email: talebira@ualberta.ca

INTRODUCTION

We present a low-cost system for markerless body part interaction detection in tele-rehabilitation. We propose: (1) dynamic thresholding normalized by inter-shoulder width, and (2) temporal filtering with majority voting. Evaluation across two camera distances (1m, 2m) shows dynamic thresholding is the primary performance driver, achieving 96% F1-score during contact phases. Existing tele-rehabilitation systems are either expensive or lack robustness for in-home use. While MediaPipe [1] provides pose data for upper limb tracking [2] and gesture recognition [3], it cannot interpret therapeutic interactions like self-touch under variable conditions. We address two key challenges: detecting contact across camera distances and filtering spurious detections from rapid movements, using only a single RGB camera.

MATERIALS AND METHODS

The system processes video frames through 4 stages: Landmark Processing: MediaPipe [1] extracts 33 pose and 21 hand landmarks per hand as 2D pixel coordinates.

1. Dynamic Thresholding and Disambiguation: We use a Dynamic threshold $d^{(k)}_{thresh} = R_k \times \left|\left|p_{ls} - p_{rs}\right|\right|_2$ where $||p_{ls} - p_{rs}||_2$ is inter-shoulder width and (R_k) is a predefined ratio. A pairwise distance matrix finds minimum distance between touching and target parts. Winner-takes-all disambiguation considers only the closest landmark pair per hand. Touch registered if distance < d^{k}_{thresh}

2. Temporal Filtering: We maintain a time-windowed queue of recent positive detections, allowing the system to track multiple concurrent interactions by confirming touches that exceed a 50% majority threshold.

RESULTS AND DISCUSSION

We evaluated using a 2×2 experimental design: dynamic vs. fixed thresholding and temporal filtering vs. raw detection. Two participants performed 8 actions (hand-to-shoulder, hand-to-head, hand-to-hip) at two distances (close: 1m, medium: 2m). Each action included 5s pre-touch, 5s touch (ground truth during middle 3s), and 3s post-touch phases. Performance evaluated frame-by-frame using Precision, Recall, and F1-score across the experimental runs. Results show: Fixed+No Filter (Precision: 0.979, Recall: 0.797, F1: 0.879), Fixed+Filter (0.983, 0.785, 0.873), Dynamic+No Filter (0.981, 0.944, 0.962), and Dynamic+Filter (0.985, 0.935, 0.960). Dynamic thresholding significantly outperforms fixed thresholding, achieving F1-scores of 0.96 vs. 0.87.

While temporal filtering shows minimal impact on single-touch performance during contact phases, it proves essential for simultaneous touch detection. The sliding window approach enables robust multi-touch capability by maintaining separate detection queues for each body part pair, allowing the system to track concurrent interactions that would be missed by instantaneous detection. This temporal aggregation reduces spurious detections during rapid movements while preserving genuine simultaneous touches. Fixed thresholding shows high precision (0.98) but lower recall (0.79-0.80), while dynamic thresholding balances both metrics effectively. The system demonstrates excellent consistency with low inter-run variability (std < 0.03).

CONCLUSIONS

We presented a low-cost system for self-touch detection with dynamic thresholding and temporal filtering. The 2×2 experimental design revealed that dynamic thresholding is the primary performance driver, achieving 96% F1-score during contact phases compared to 87% for fixed thresholding. Temporal filtering proves essential for simultaneous multi-touch detection through its sliding window approach, enabling robust tracking of concurrent interactions while maintaining high precision (0.985). The system demonstrates excellent reproducibility with low inter-run variability. Applications include remote fitness coaching, sports training, and ergonomic monitoring.

REFERENCES

[1] Lugaresi C et al. arXiv:1906.08172, 2019.
[2] Wagh V et al. JMIR Form Res 8: e56682, 2024.
[3] Jo BJ et al. Ingénierie des Systèmes d'Information 28: 633-8, 2023.