Hand motion estimation in the wild via IMUs and gaze-tracking: Can an efficient linear model suffice?

TECHNION Israel Institute of Technology

Eitan Gerber, Nili E. Krausz

Neuro-Robotics and Bionic Limbs (eNaBLe) Lab

Technion - IIT

Introduction

- Amputees are **unable** to coordinate multi-joint robotic prostheses, in part due to limited available inputs.
- Previous work has shown correlation between gaze, sEMG and hand motion [1].
- We propose a **multi-modal sensor fusion model** as an endpoint predictor for a prosthetic arm controller.
- This work may also have implications for motion retargeting to address **sim2real** for humanoids and other robotics applications [2].

Methods

- We used Meta's open-access Project Nymeria dataset [3].
- Used 9 subjects, 8 for training + 1 for testing.
- Trained a series of **linear regressors (LR)** using MATLAB.
- Trained a modified **UNet** in PyTorch, with a GRU layer at the bottom.
- Predictors: **Upper-arm coordinates**, **shoulder angles**, **gaze vector**, **gaze pitch & yaw angles**.
  - Including velocities/accelerations of arm, shoulder.
  - Upper-arm coordinates = mean of the shoulder and elbow coordinates.
- Targets: **Hand coordinates.**
- Trained the **UNet** with a loss function that penalizes Euclidean distance and differences in velocity, and **minimizes jerk**.
- Compared the accuracy of the **LR** when using individual predictors and when using all predictors.

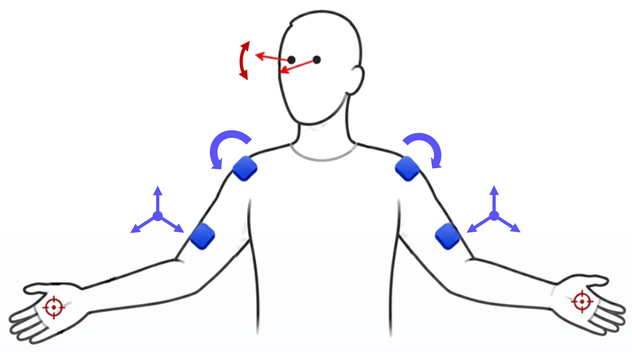


Fig 1. Illustration of the predictors and targets used.

Results

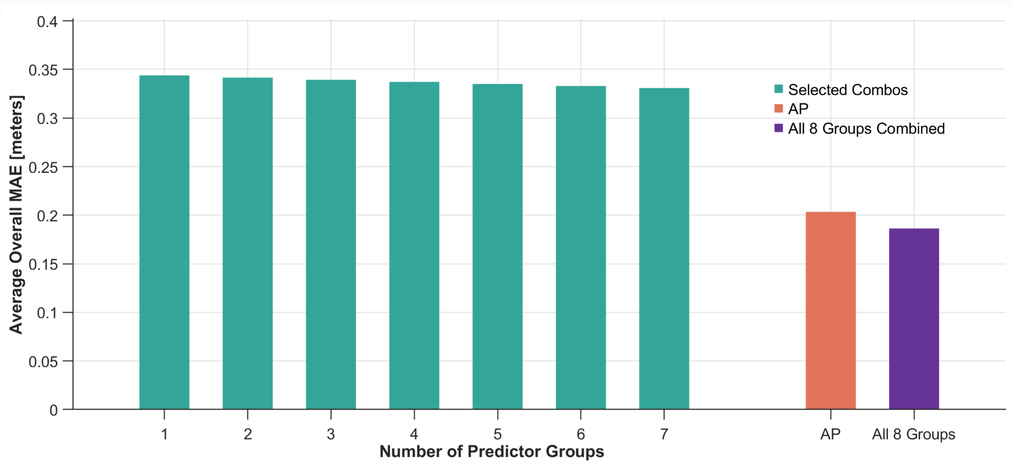
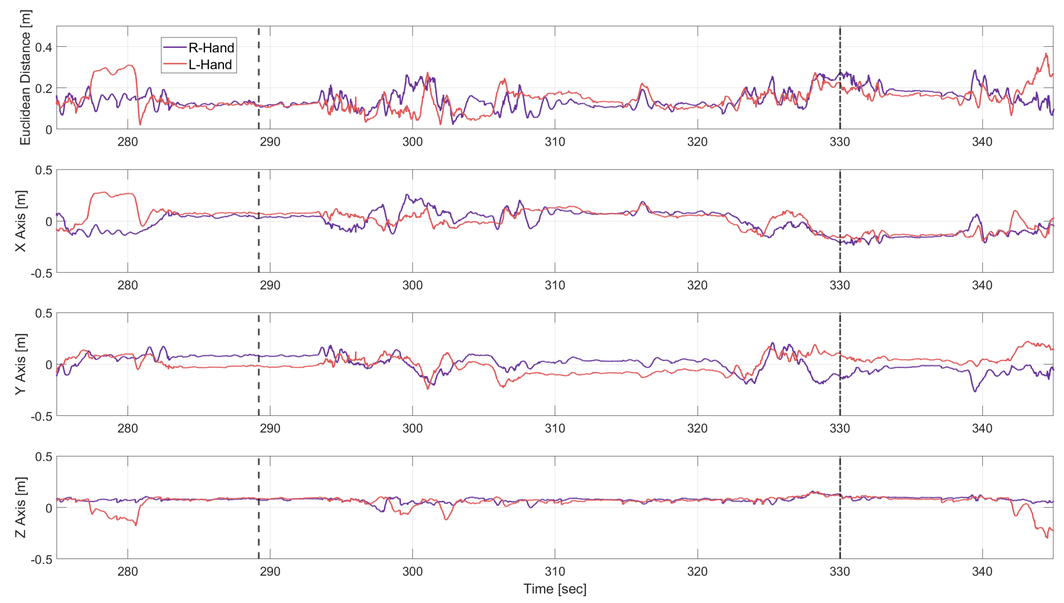
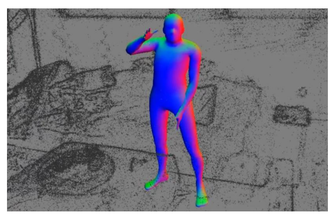


Fig 2. Overall average 3D MAE produced by LR with all possible combinations of predictor groups other than the arm position (AP) vs. the number of predictor groups used together. On the right-hand side of the graph we show mean 3D MAE using AP group alone, vs. all 8 predictor groups.

(a)



(b)



(c)

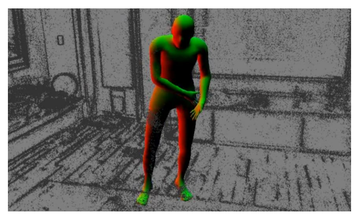


Fig 3. Distance graphs for a representative portion of the test data showing segments with higher and lower LR prediction accuracy **(a)**, corresponding to the hands being used to hang a streamer **(b)** and the hands being used to bunch-up a cloth **(c)**.

- **LR** predictor achieves 3D MAE of $\mathbf{18.6 \pm 1.9}$ **cm** over 8 train / 1 test subjects.
- **UNet** predictor achieves 3D MAE of $\mathbf{14.2 \pm 8.0}$ **cm** over 8 train / 1 test subjects.
- Both predictors showed strong generalizability between tasks and subjects.
- **Training** and **inference** times differed significantly:

	Training Time	**Mean Inference Time**
LR predictor	minutes	$0.54 \cdot 10^{-3}[ms]$
Unet predictor	>24 hours	$1.49[ms]$

Table I: Training and inference times for the tested predictors.

Linear Regression provided surprising prediction accuracy for ADLs compared to deep learning

Conclusions and Future Work

- This work demonstrates the potential for computationally efficient prediction.
- Acquire full dataset combining MoCap, IMU, sEMG, and gaze-tracking data recorded during simulated ADLs – see Fig. 4.
- Identify relevant **synergies** between sensor modalities.
- Create an efficient **real-time endpoint predictor and controller**.
- Control a robotic hand and Franka arm to demonstrate real-time ***prediction of user intent***.
- Test controller with transhumeral amputees and custom prosthesis.

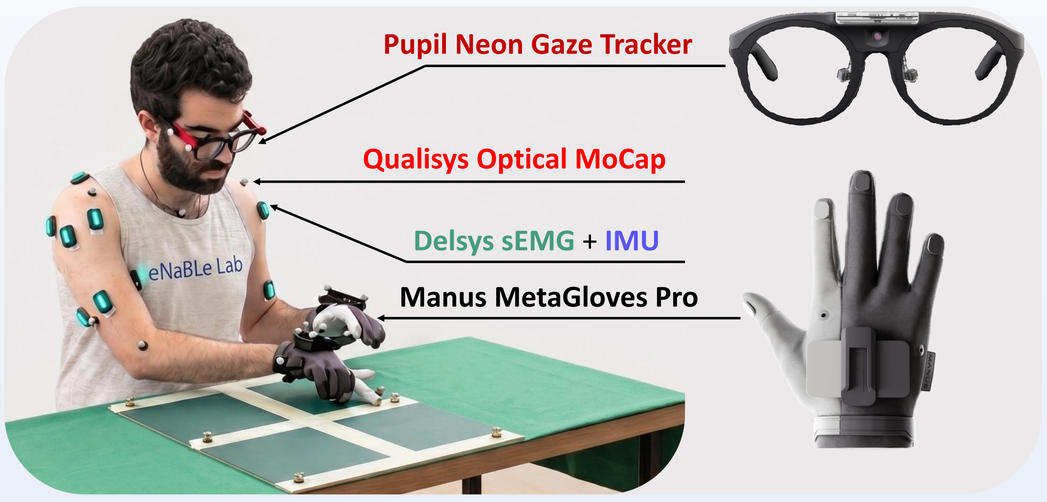


Fig 4. Illustration of the dataset acquisition setup and sensors.

[1] N. E. Krausz, D. Lamotte, I. Batzianoulis, L. J. Hargrove, S. Micera, and A. Billard, "Intent Prediction Based on Biomechanical Coordination of EMG and Vision-Filtered Gaze for End-Point Control of an Arm Prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1471–1480, Jun. 2020, doi: 10.1109/TNSRE.2020.2992885.
[2] E. Aljalbout et al., "The reality gap in robotics: Challenges, solutions, and best practices," Annu. Rev. Control. Robot. Auton. Syst., vol. 9, pp. 403–432, May 2026.
[3] L. Ma et al., "Nymeria: A massive collection of multimodal egocentric daily motion in the wild," in Proc. ECCV, 2025, pp. 445–465.

This research is funded by the Israel Ministry of Innovation, Science and Technology Grant #0007458