Sensor Fusion System for Improving Motion Amount Quantification within CV-Enabled Worker Analysis

Ethan Chang, Computer Systems Engineering
Shenghan Guo, Assistant Professor
School of Manufacturing Systems and Networks

Research Question

How can computer-vision-enabled motion amount quantification be improved without changing the computer-vision model in a way that is simple, inexpensive and effective?

Approach Brainstorming

1. Optimize Data Preprocessing and Feature Engineering
   a) Ensuring data diversity and consistency is a user effort
   b) Optimizations should happen during the training process
2. Using Ensemble Methods
   a) Combining multiple outputs is computationally taxing
   b) Aggregation techniques are both simple and reliable
3. Incorporate Richer Contextual Information
   a) Combining multiple outputs is computationally taxing
   b) May perform better than simple aggregation due to different types of data
   c) More expensive due to requiring different sensor types
4. Perform Data Post-Processing
   a) Additional filtering and smoothing will help reduce noise and improve stability of motion quantification

Chosen Approach + Reasoning

Elimination:

#1 changes the model and requires user effort ▸ Rejected
#2 is costly but simple and reliable ▸ Accepted
#3 is very costly but performs well ▸ Accepted
#4 has no downsides except developer investment ▸ Accepted

Choosing between #2 and #3:

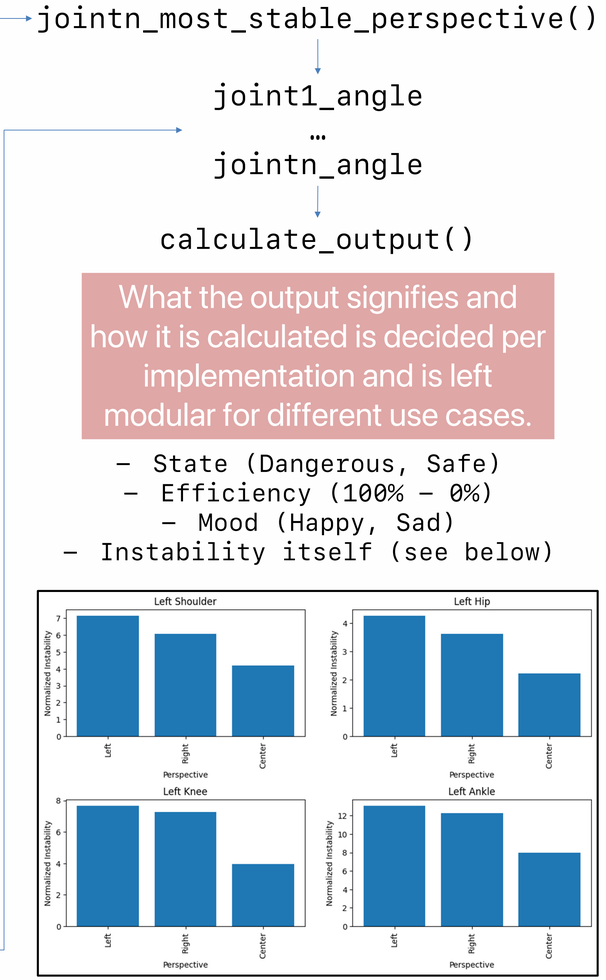
Considering the context, worker analysis for general motion amount quantification: Soft real-time system.

If the system performs reasonably well, cost is of more importance to the target user than an improvement to output accuracy. Thus, choose #2.

Final Choice:

Use Ensemble Method (homogeneous sensor fusion) with Data Post-Processing (filtering)

Current Progress



`instability` represents the **cumulative Euclidean distance traveled by a specified joint at a specified frame**. This way of measuring accuracy was chosen for (1) its simplicity and (2) because erratic movement of landmarks in the video feeds were the most common symptom of bad output.

Project Objective

Increasing the accuracy of computer-vision-enabled motion amount quantification in a simple and intuitive way, with little to no user investment.

Future Directions

- Optimizing the performance of the sensor fusion system so lower-end devices can reliably perform computations in real-time
- Improving the instability metric to something better than Euclidean distance
- Creating other presets for `calculate_output()` other than instability and state, with accompanying control system
- Exploring model architecture tuning and other techniques that are not in the scope of this project

Acknowledgements

The pose estimation model I used to process landmarks is a MediaPipe solution that you can find at this link - *https://developers.google.com/mediapipe/solutions*

I would like to thank my mentor, Shenghan Guo, and my graduate student advisor, Pius Gyamenah, for their help on this project.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University