Visual Reconstruction of Human Activities from data generated using Inertial Motion sensors

School of Computing, Newcastle University
Koustav Barik, Dr. Jacek Cala

Motivation

For **Human Activity Recognition** (HAR) research, **generating a comprehensive dataset** with diverse activities & subject demographics is often a **laborious process** needing a huge amount of time, money, & effort.

To fix this problem, **generative models** can be used to **synthetically produce data**, but there is **no method to visually reconstruct** sensor data **& validate its consistency** with activity labels. This visual reconstruction can aid in **recognition of bad training data** for generative networks & provide a metric to **assess the reliability of the synthetically generated data**.

Objective

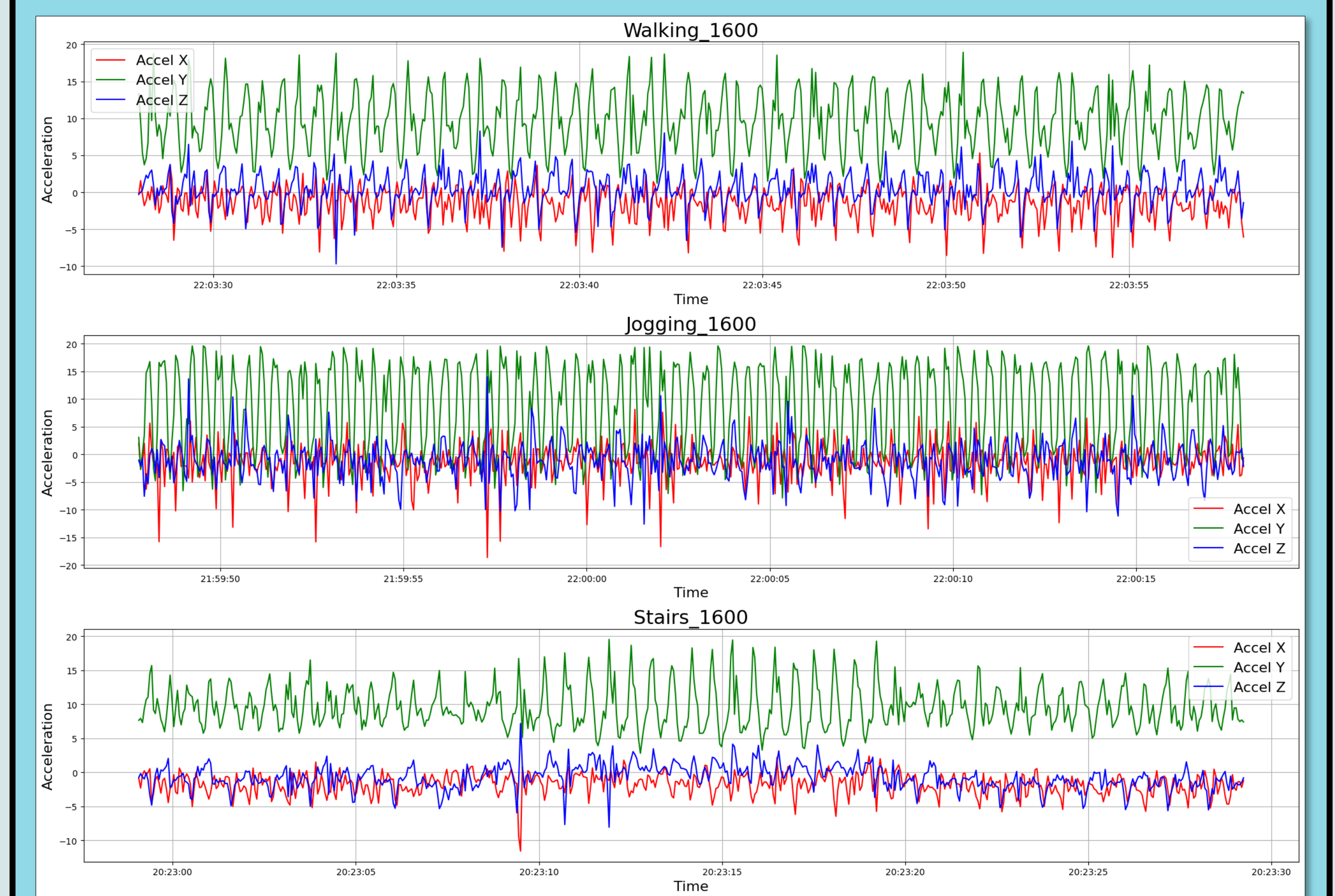
- This project aims to use motion tracking data to **visually reconstruct human movement** in 3 dimensions with **6 degrees of freedom**, with the intention of manually classifying human activities & validating them against a set of ground truth labels.

Dataset & Framework

- **WISDM dataset [1]** was choosen
  - as it offers **ample training data** for generative networks.
  - Its **raw data format** allows for customized preprocessing, correction of inconsistencies & real-world data scale for visual reconstruction.
  - It includes recordings from **both accelerometer** & **gyroscope sensors** of smartphones.

Out of 17 available activities, we chose the below **3 activities**

Activity Classes	Walking	Jogging	Stairs
Observations	32.60%	33.01%	34.39%

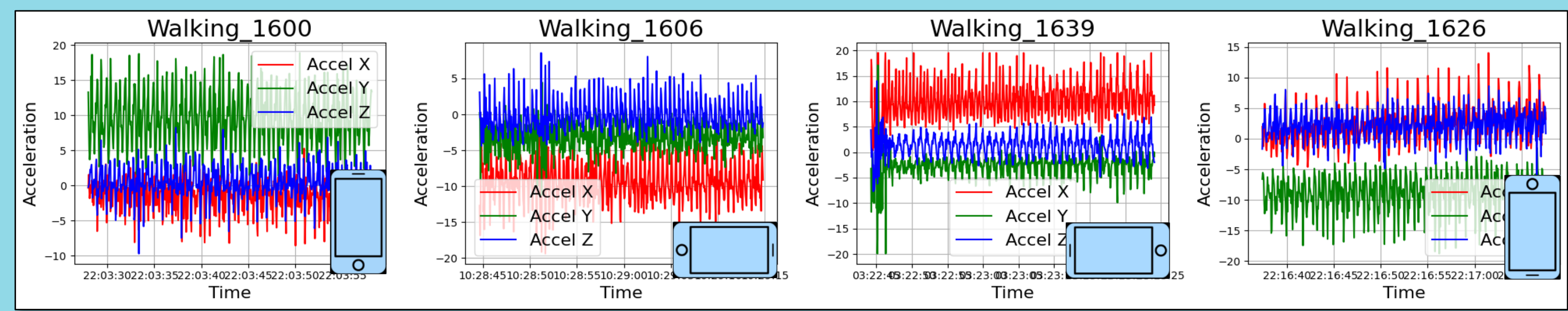


- After exploring multiple frameworks, assessing their limitations & fit with respect to our requirements, we chose the **Unity [2]** game engine for visual reconstruction of human activity data. It provided us with a platform to use both displacement & orientation data to reconstruct the motion data.

Methodology

Data Issues fixed

- **Phone Orientation issues**



We fixed this by shifting the signals, i.e., adding two times the absolute average of the amplitude and exchange X and Y axes if the average amplitude of the X-axis was more significant than the average amplitude of the Y-axis [3].

- **Sampling issues**

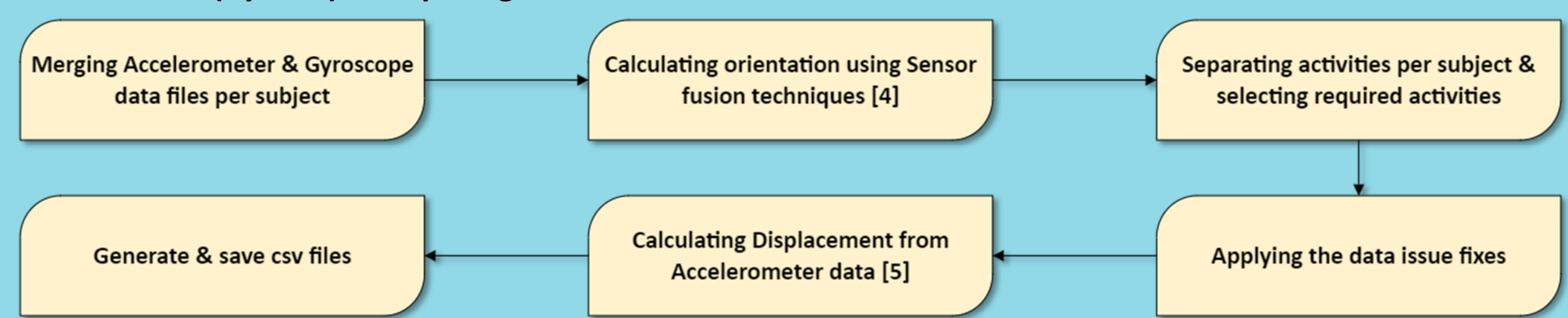
Data collected on different smartphones resulted in different sampling frequencies. For example, No. of sample points: ~**3600 points** for subjects 1600, 1602, 1604, ~**4500 points** for subjects 1601, 1647, ~**14280 points** for subject 1629, etc. We fixed this by resampling the frequency to 20 Hz [3].

- **Bad data removal**

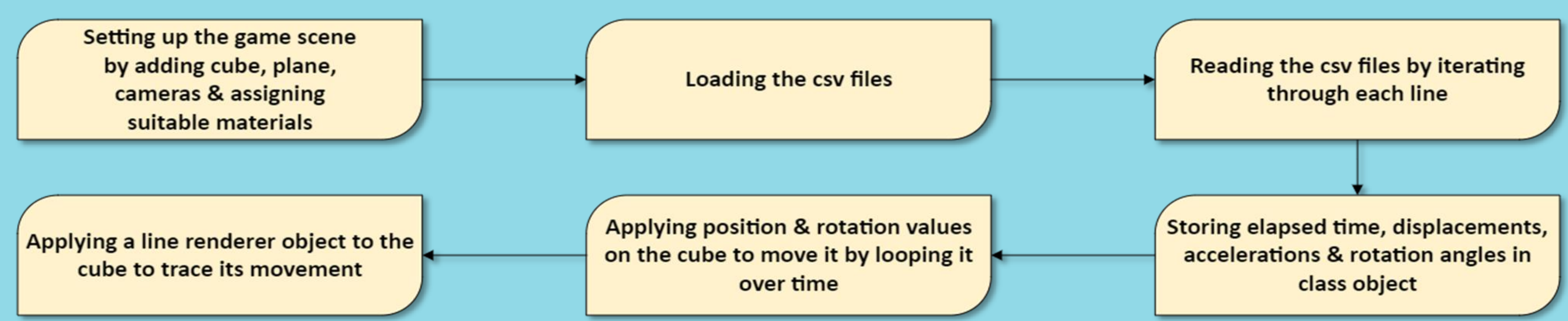
For some activities, in case of some subjects, there is not enough data. For e.g., in case of Jogging, subjects 1614, 1641, 1642, 1644, 1645, 1646., have **less than 300 data points**. We removed them.

Visual Reconstruction process

- **VS Code (Python) – Preparing the csv files:**



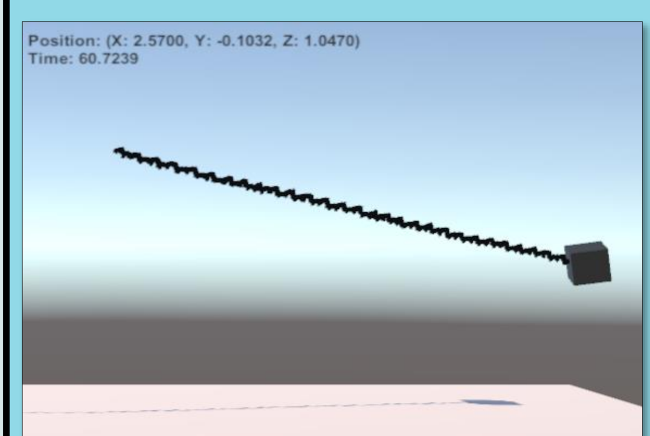
- **Unity (C# Script) – Processing the csv files:**



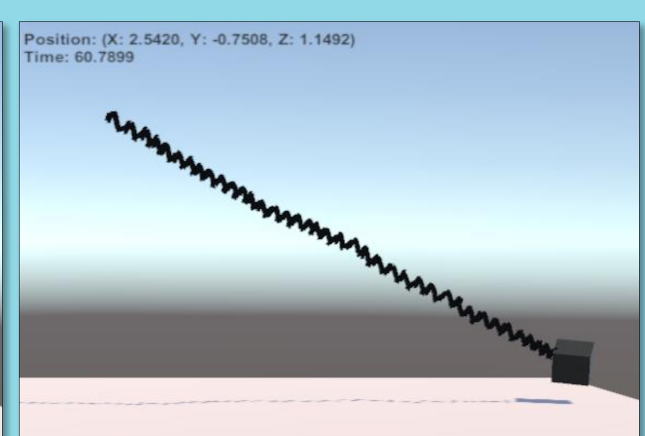
Visual Reconstruction outputs

- For Jogging, the cube portraying the phone kept in the subject's hip pocket showed more frequent ups & downs as compared to Walking & Stairs.
- The distance travelled was more in the case of Jogging than Walking & Stairs over a fixed period.
- The Stairs movement showed a change in pace & orientation in fixed intervals which was a result of the landing area after a flight of stairs.

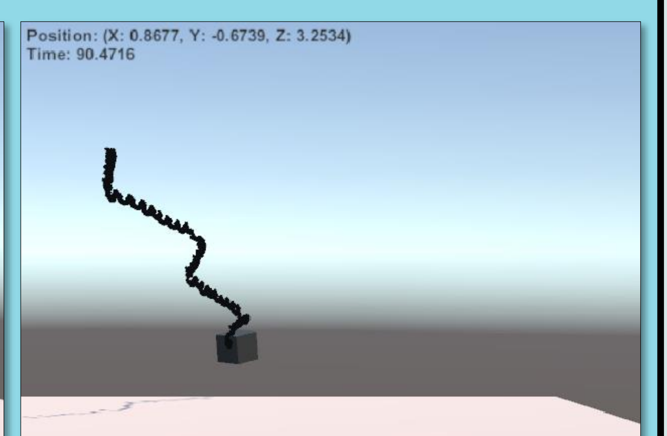
Walking



Jogging



Stairs



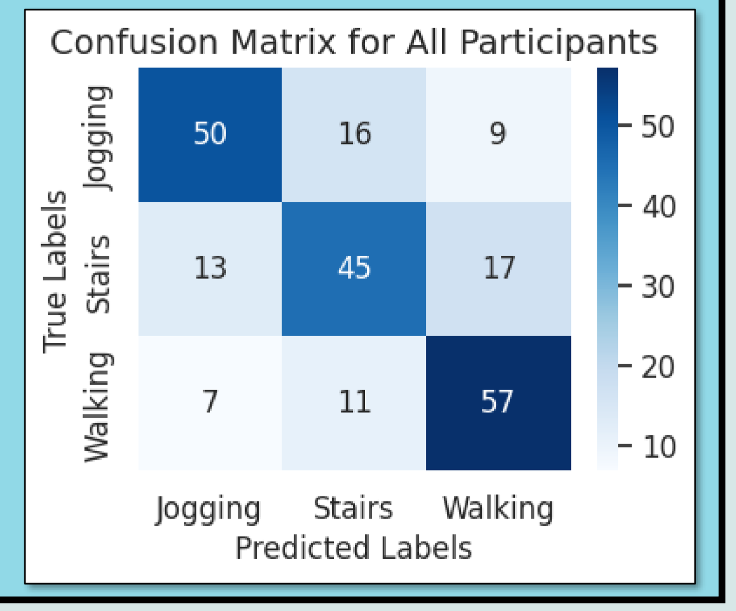
Results

- **Result Collection**
  - 15 test cases were created – 5 sets for 3 activities.
  - A form was sent out with unnamed labels for surveyors to fill out.
- **Metrics**
- **Percentage of Activities Predicted Correctly:**

Walking – 76%	Jogging – 66.67%	Stairs – 60%

- **Classification Report & Confusion Matrix:**

	Precision	Recall	F1-score
Walking	0.69	0.76	0.72
Jogging	0.71	0.67	0.69
Stairs	0.63	0.60	0.61
Overall Accuracy	0.68		



Discussion

- **Evaluation**
- The current setup used to visually reconstruct human activities showed that the outputs can be identified by humans with some accuracy.
- Its easier to compare reconstructed visuals and label them than looking at an isolated visual when viewing it without prior context.
- Its easier to identify Walking activity compared to other two, since Stairs and Walking was almost similar in case of some reconstructions, and some Jogging reconstructions showed deviation away from a straight line of motion.

- **Limitations**
- The reconstructions showed a sinking effect in the Y-axis which might be due to accumulation of errors while converting acceleration to displacement.
- The current setup is not useful in case of multi sensor data obtained from sensors placed in multiple body locations of a human subject.

- **Future Work**

We look to address the limitations discussed above by exploring filtering techniques and better sensor fusion algorithms. The multi sensor approach can be looked at using a humanoid model in Unity and applying forces in different joints. While this can be a starting point for future work, it needs rigorous experimentation to make a multi sensor visual reconstruction model work due to error accumulation and calibration issues.

References

- [1] G. M. Weiss, K. Yoneda, and T. Hayajneh. "Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living". In: IEEE Access 7 (2019), pp. 133190–133202.
- [2] A. Juliani et al. Unity: A General Platform for Intelligent Agents. 2020. arXiv: https://arxiv.org/abs/1809.02627
- [3] M. Heydarian and T. E. Doyle. "rWISDM: Repaired WISDM, a Public Dataset for Human Activity Recognition". In: arXiv preprint arXiv:2305.10222 (2023).
- [4] Reference listTechnologies, x-io (2024). xioTechnologies/Fusion. [online] GitHub. Available at: https://github.com/xioTechnologies/Fusion?tab=readme-ov-file [Accessed 31 Jul. 2024].
- [5] A. Blake, G. Winstanley, and W. Wilkinson. "Deriving Displacement from 3-Axis Accelerometers". In: Computer Games, Multimedia & Allied Technology. 2009.

CONTACT : k.barik2@newcastle.ac.uk | koustavbrk@gmail.com