Supplementary Material
Egocentric Activity Recognition on a Budget
ID: 3516

1 Detailed Results

Figure 1 shows the confusion matrices on the multimodal dataset for motion and vision individual predictions and also for the combined usage through the policy framework.

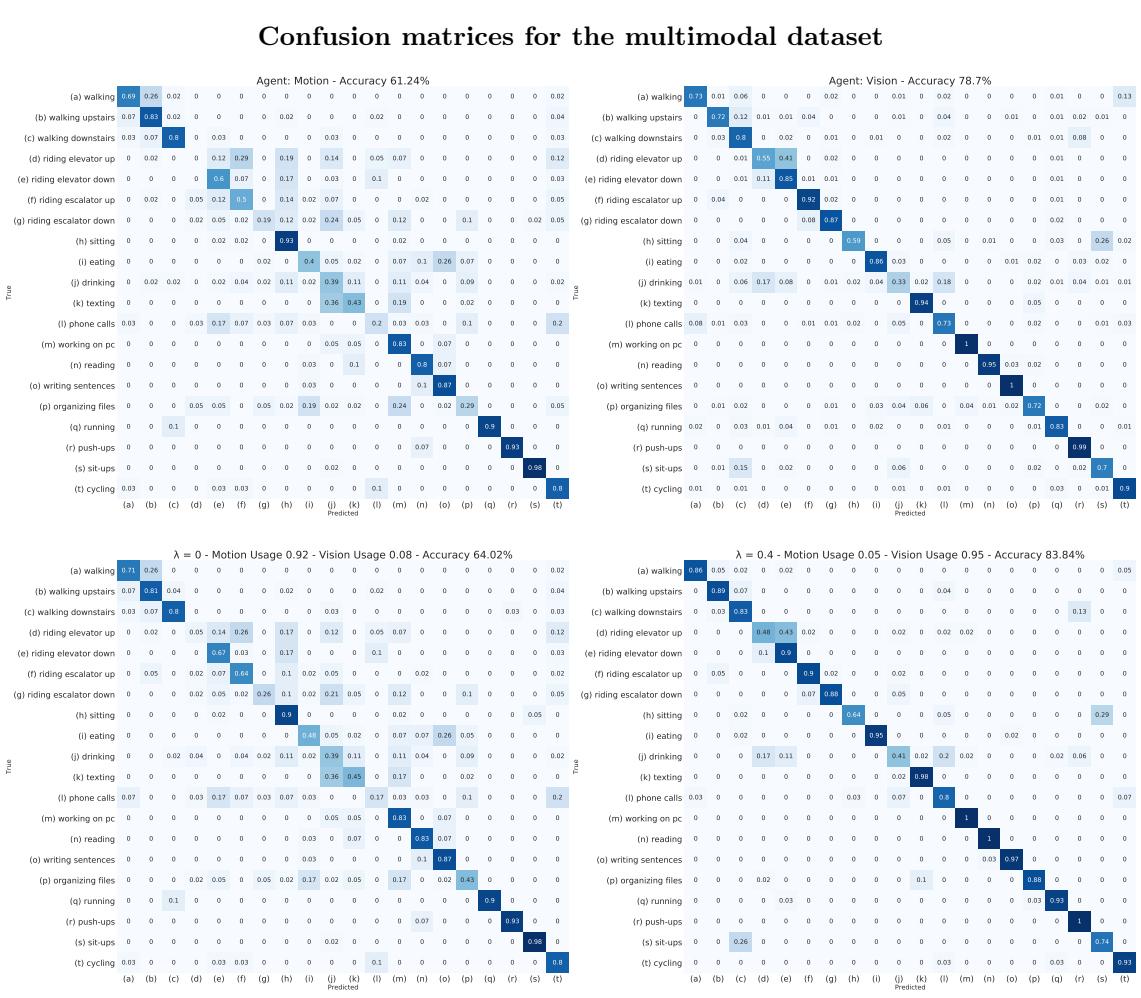

Figure 1: Confusion matrices shows better accuracy/energy tradeoff when using an optimized policy

2 DataEgo Novel Dataset

We created the novel egocentric dataset DataEgo. We make this dataset publicly available online. This dataset is temporarily unavailable for blind review purposes.

DataEgo contains sequences of natural activities, totalizing approximately 4 hours of recording. This dataset considers 20 classes organized in a taxonomy of 5 groups as can be seen in Table 1. Thus, it is possible to customize the classification by considering the 20 activities or their higher level abstraction which contains 5 groups.

Activity	**Category**
Walking	Mobility
Walking Down/upstairs	
Riding Elevators	
Riding Escalators	
Working on PC	Office Work
Reading	
Writing	
Chopping Food	Kitchen Related
Cooking on Stove	
Washing Dishes	
Running	Exercising
Doing Push ups	
Doing Sit ups	
Cycling	
Eating and Drinking	General
Talking with People	
Browsing Mobile Phone	
Lying Down	
Watching TV	
Brushing Teeth	

Table 1: Set of Activities in DataEgo

2.1 Data Collection

The data in our dataset was captured with the Vuzix M300 Smart Glasses shown in Figure 2. This device allows us to record egocentric video and accelerometer/gyroscope information. Our dataset was captured by different volunteers in order to introduce into the dataset the natural variability of an activity being performed by several individuals. Figure 3 illustrates the collection process.

Furthermore, our dataset is fully tagged. Initially, the tagging process was partially performed by tapping twice a panel located on the side of the device which indicates a change of activity. Nevertheless, we realized that this fact was requiring an extra effort from the volunteers, entailing confusion and partially affecting their natural behavior. Thus, we performed the labeling of the dataset in a post-capture step.

Figure 2: Vuzix Smart Glasses.

Figure 3: The sequence of activities was developed by different subjects.

2.2 Sample Description

Each recording is composed of a realistic sequence of 4 to 6 distinct activities totalizing 5 minutes. Figure 4 exhibits some video frames and accelerometer plots from a single recording. This recording is comprised of a sequence of 6 activities where Walking is repeated once. We can appreciate that sensors are able to not only provide better results on activities with clear movements patterns such as running or walking but they are also able to replace vision methods on activities where images are not properly captured (e.g. Rainy and Dark environments). Using both sources of data interchangeably improves both energy consumption and overall accuracy of our framework.

2.3 Intra Class Variability

A real life activity can be performed in a wide range of conditions. Aiming to capture these conditions in our dataset, the samples were recorded in a wide variety of scenes. Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 represent a set of examples of the captured data for each of our 20 activities. In the top part of these Figures, we can appreciate the visual information obtained by the egocentric camera. The plots located in the middle and bottom parts represent the information from the accelerometer and gyroscope respectively. The colors red, green, and blue represents the x, y, and z axis respectively in both, the accelerometer and the gyroscope plots. Analyzing the visual information in these Figures, it is evident the diversity of scenes for each of the activities. Also, by analyzing the sensors information, some variability can be noted for a given activity. For example, the sensors readings of the three examples of the Doing Sit Ups activity from Figure 17 exhibit a repetitive pattern. However, the length of this pattern varies considerably in the second example. Also, the y axes of the accelerometer (green) of the second and third examples have a visible repetitiveness, which contrast the first example. These facts unveil the intrinsic nature of an activity which varies according to the subject, and conditions. Each individual in a specific setting has a particular manner to perform an activity which determines its speed, and movements. Being able to capture this variabilities from real life highlights the realistic nature of our dataset.

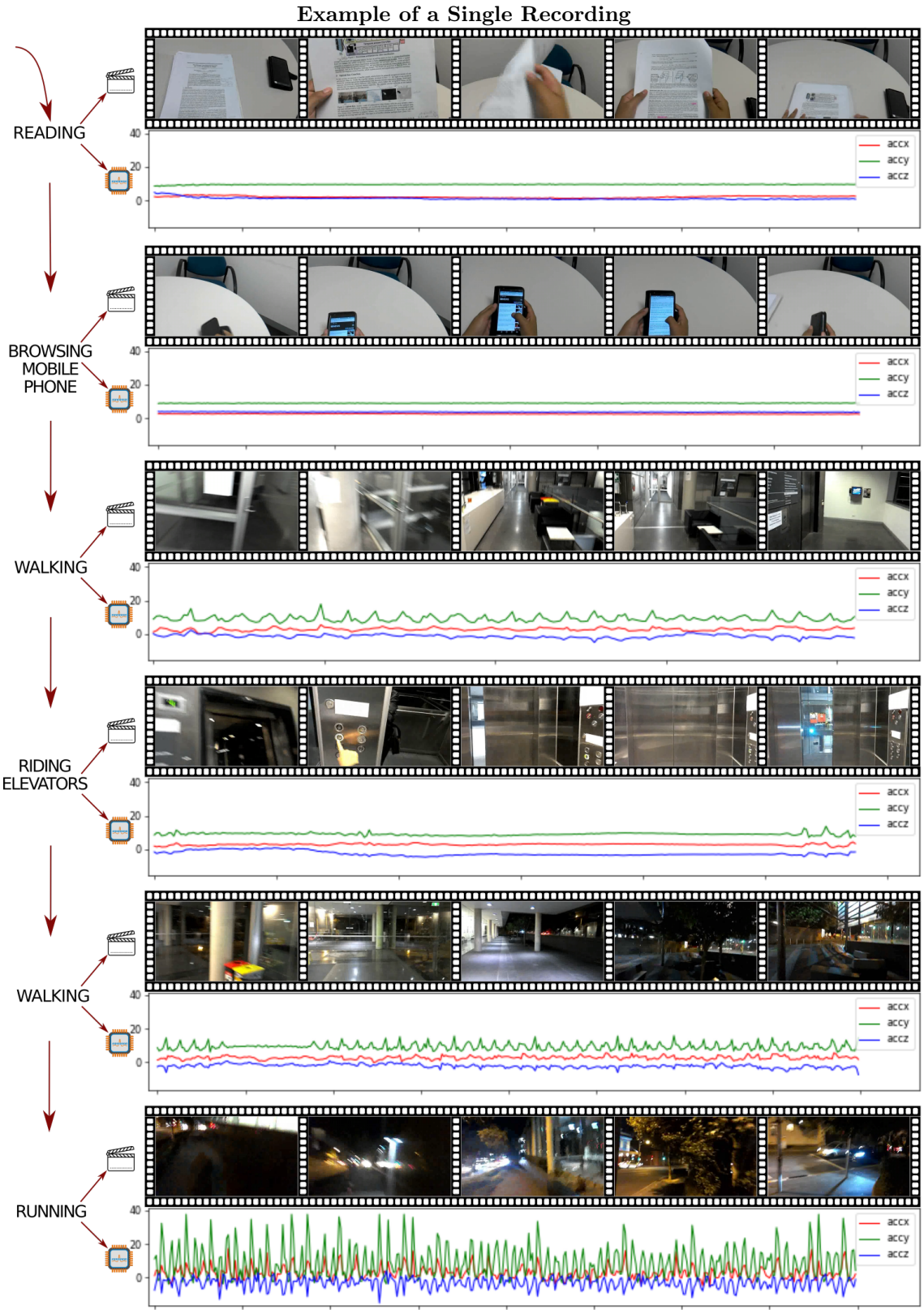

Figure 4: Example of a single recording of our dataset showing a sequence of activities.

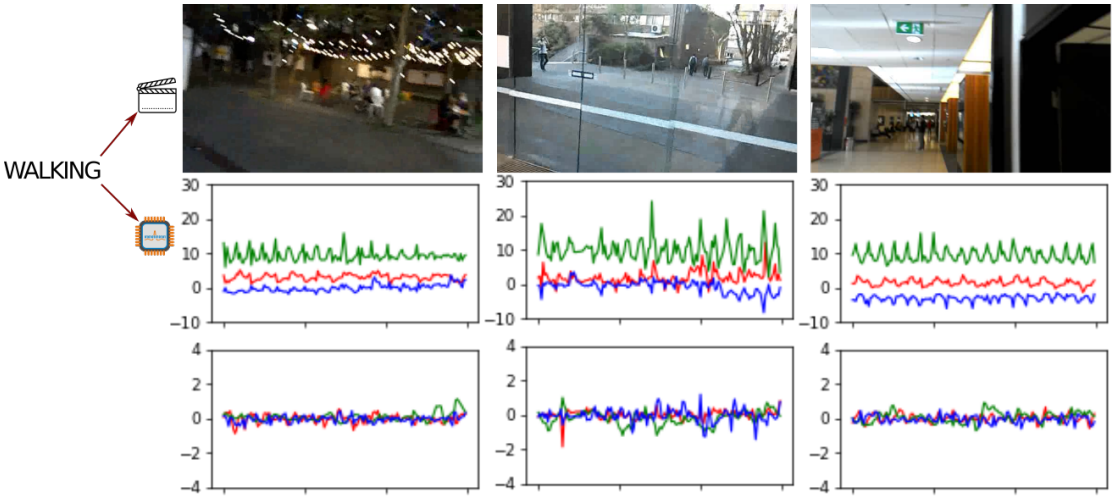

Figure 5: Activity Walking.

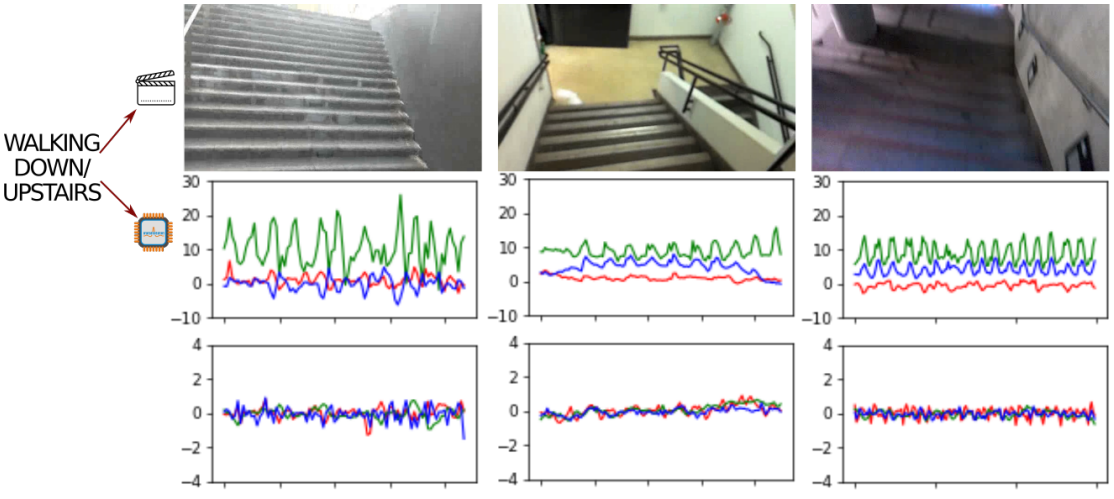

Figure 6: Activity Walking Down/Upstairs.

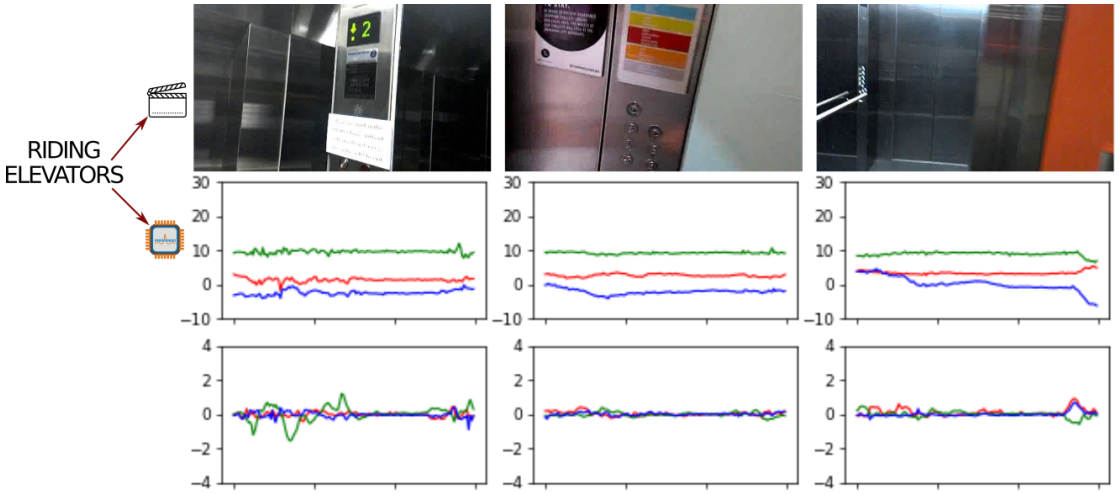

Figure 7: Activity Riding Elevators.

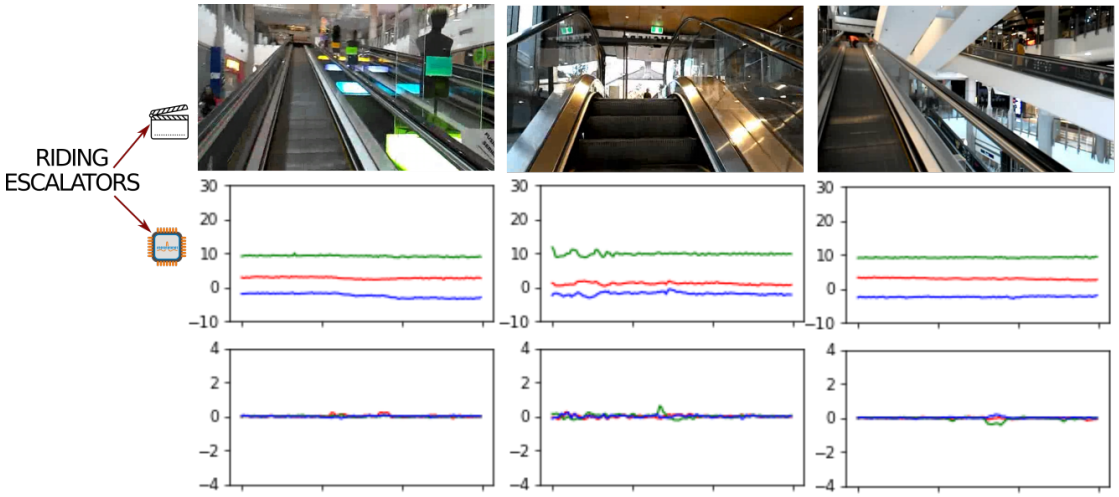

Figure 8: Activity Riding Escalators.

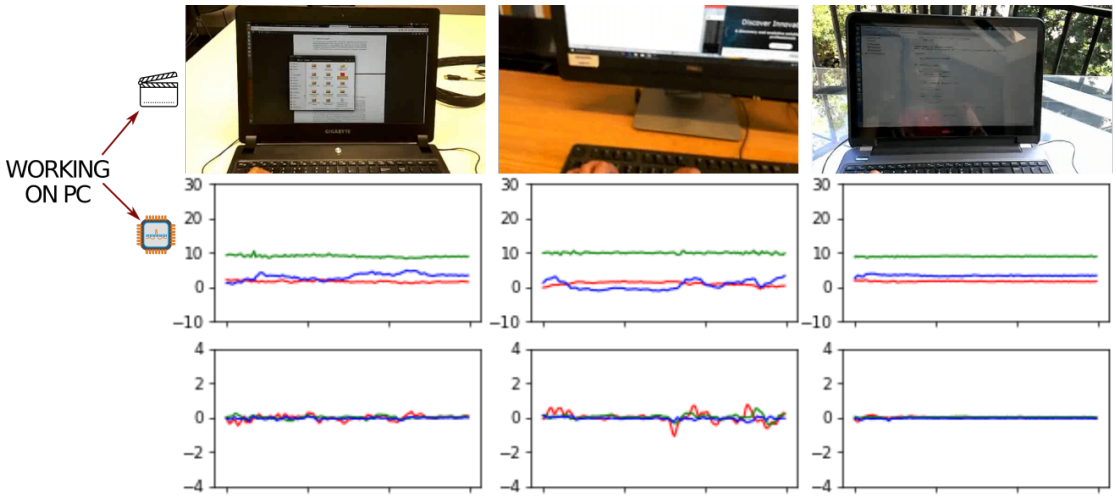

Figure 9: Activity Working on PC.

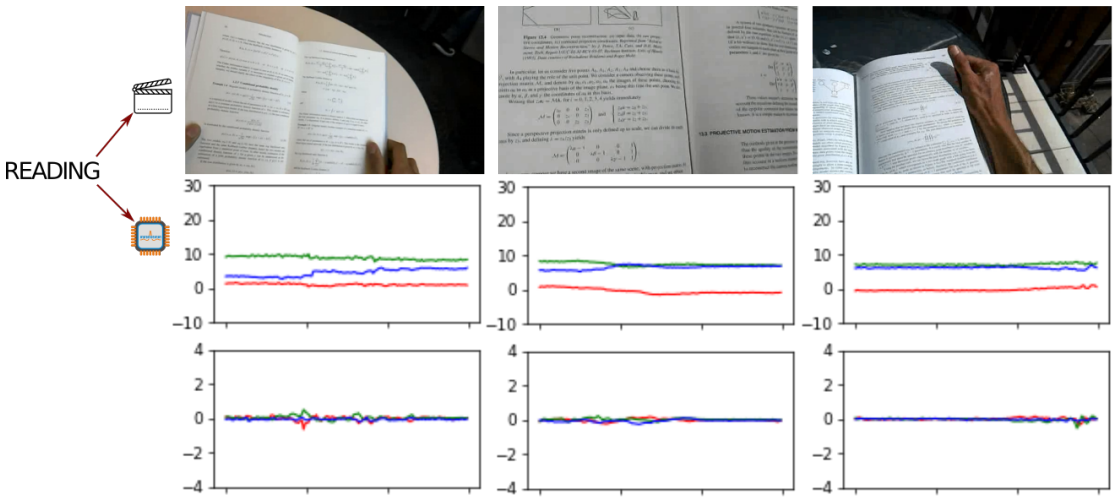

Figure 10: Activity Reading.

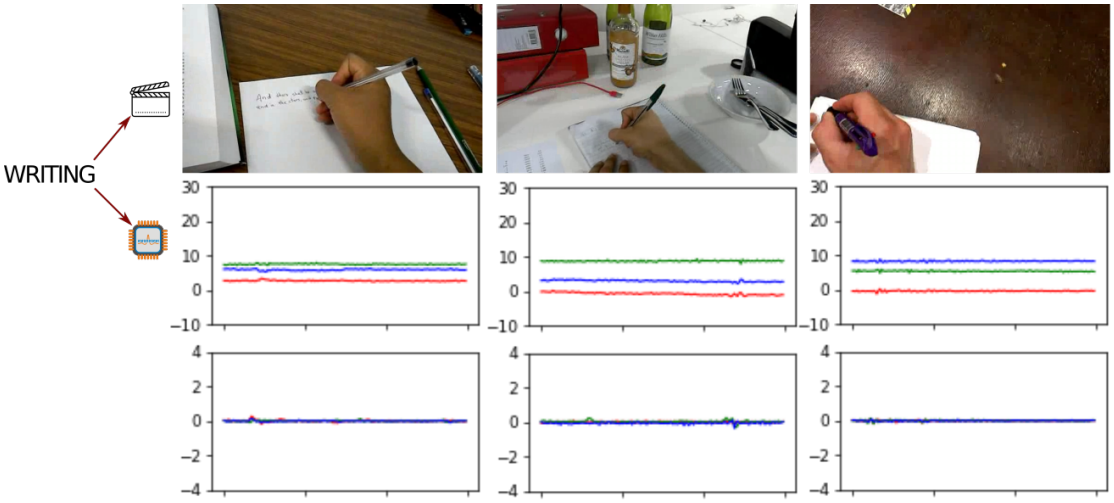

Figure 11: Activity Writing.

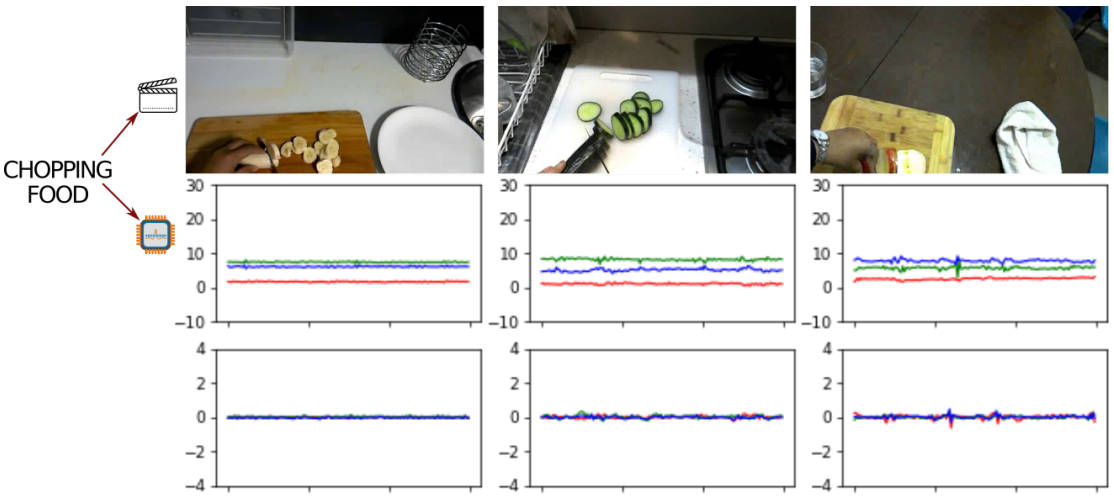

Figure 12: Activity Chopping Food.

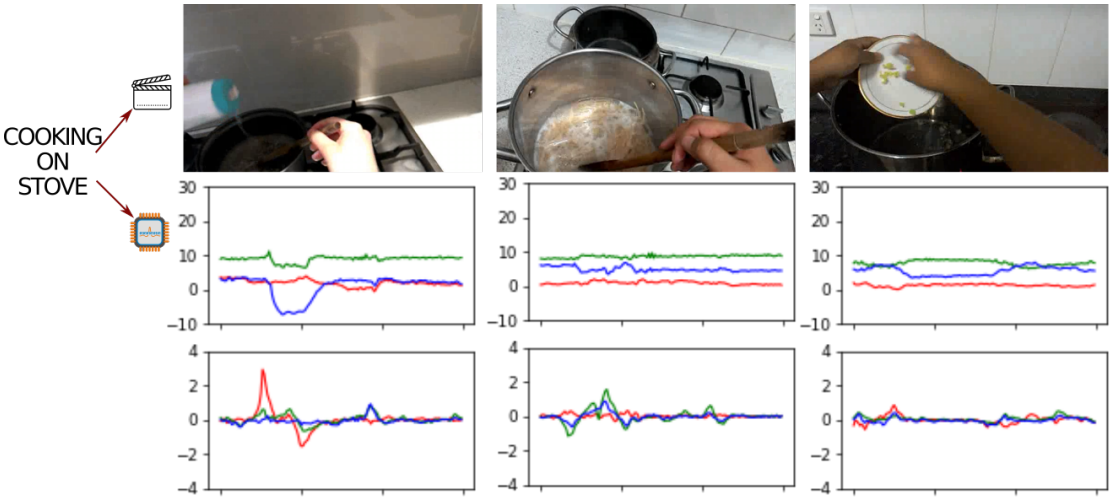

Figure 13: Activity Cooking on Stove.

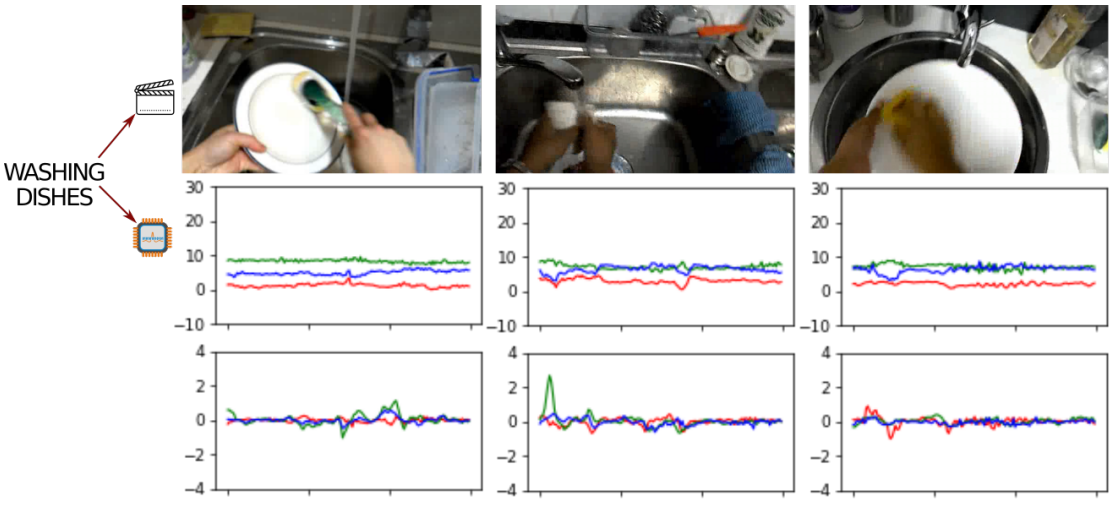

Figure 14: Activity Washing Dishes.

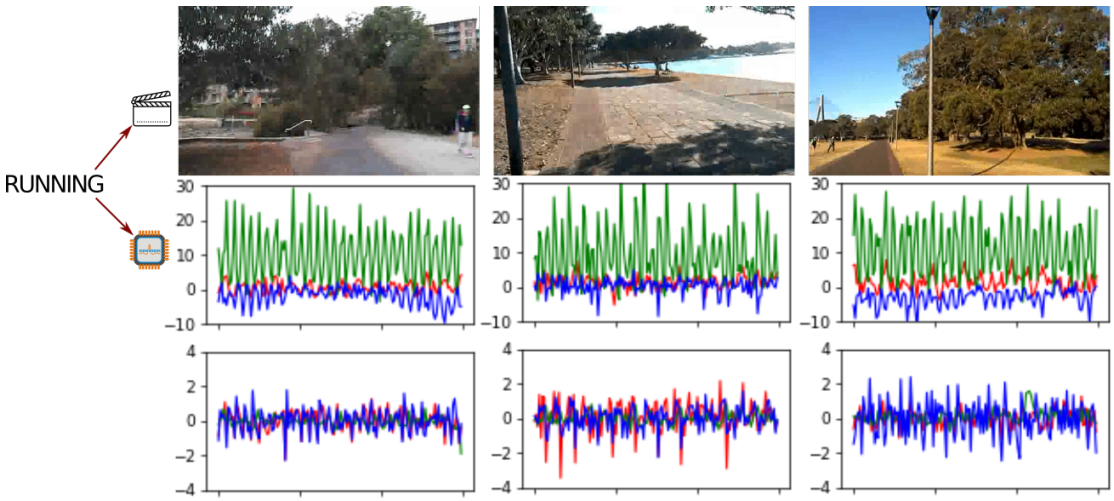

Figure 15: Activity Running.

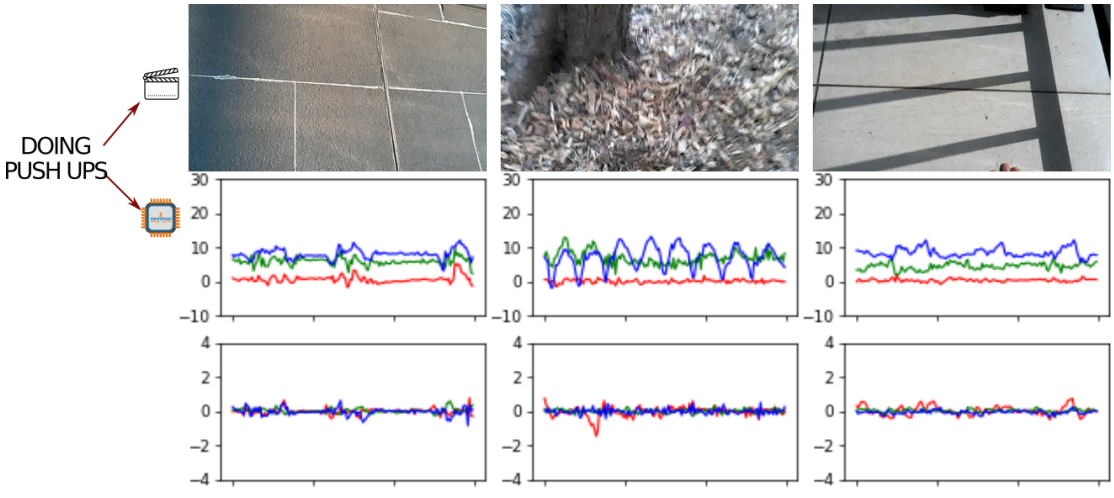

Figure 16: Activity Doing Push Ups.

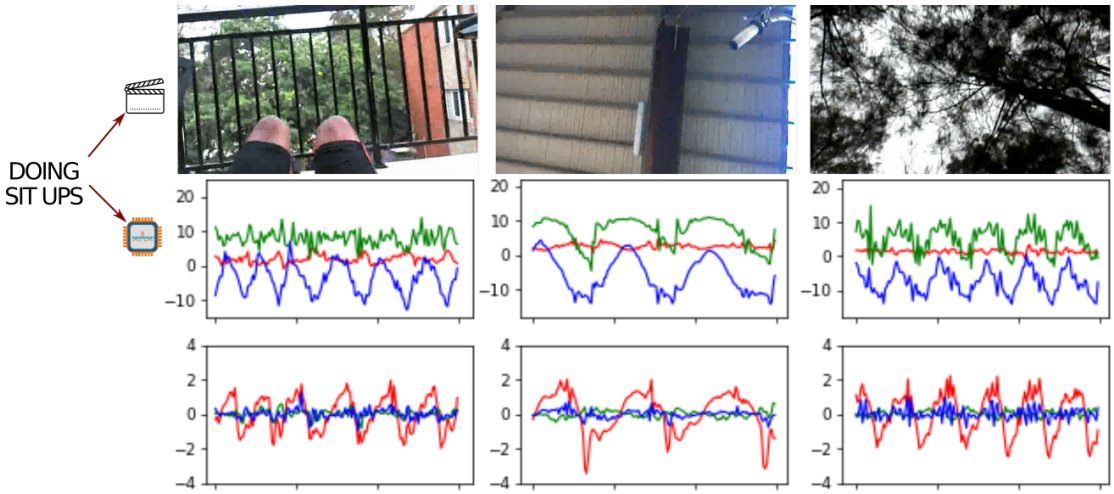

Figure 17: Activity Doing Sit Ups.

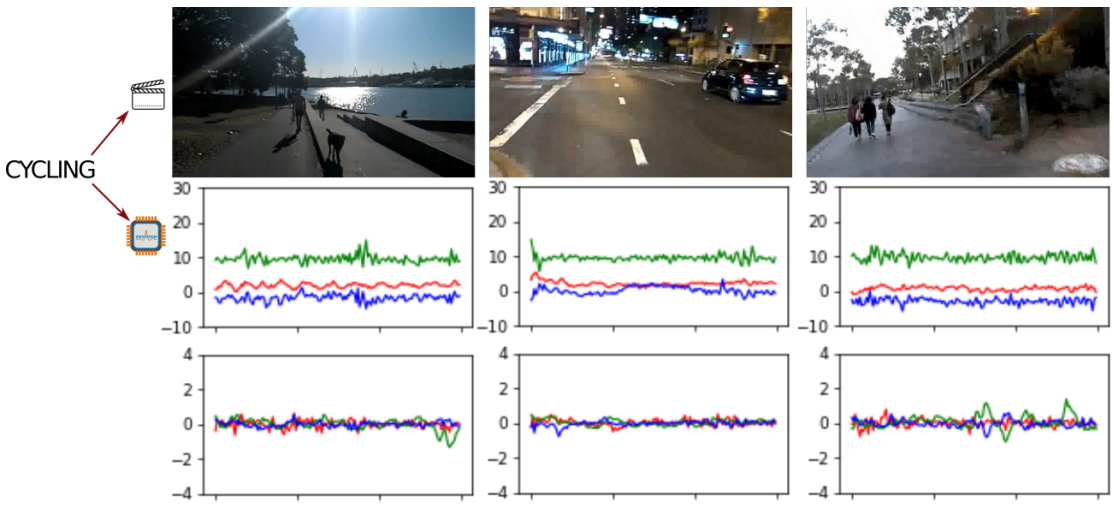

Figure 18: Activity Cycling.

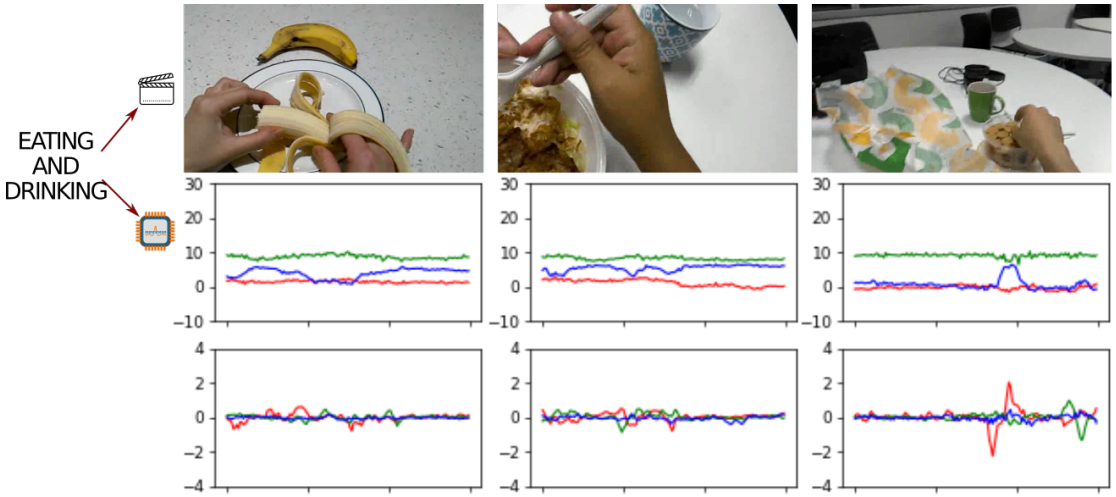

Figure 19: Activity Eating and Drinking.

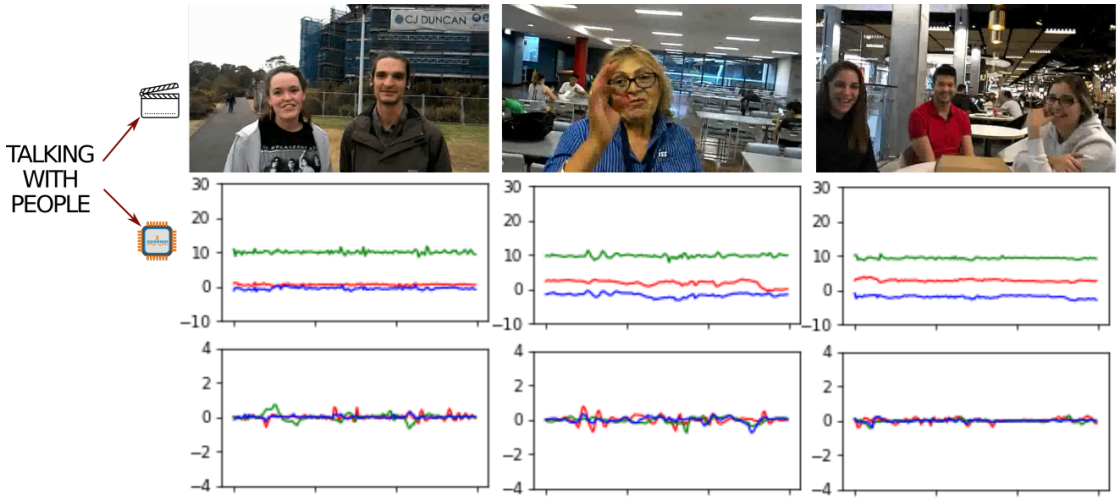

Figure 20: Activity Talking with People.

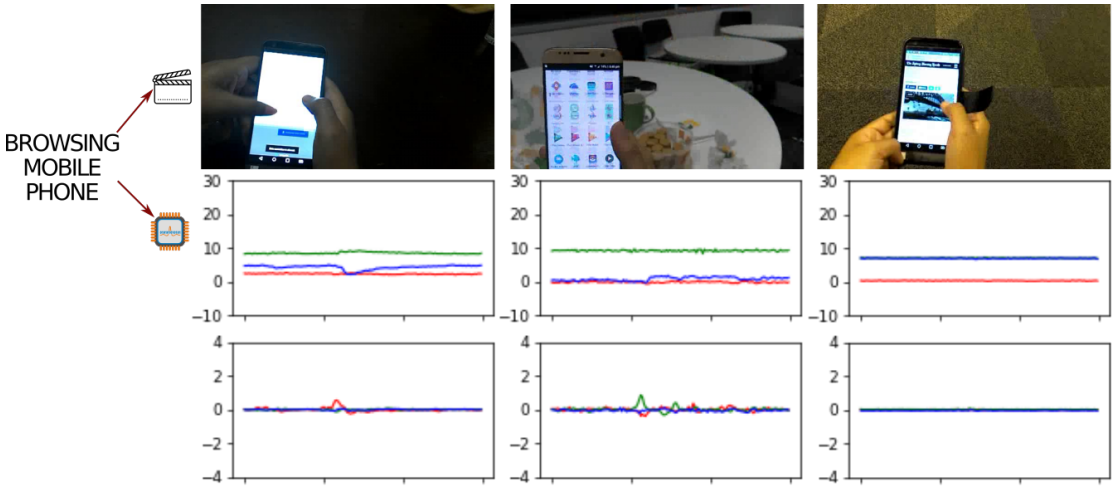

Figure 21: Activity Browsing Mobile Phone.

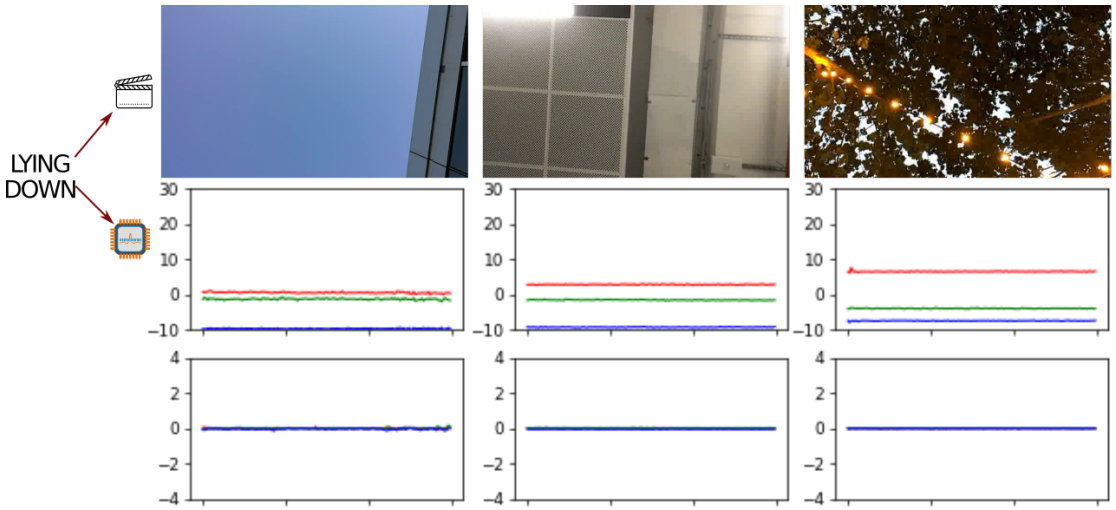

Figure 22: Activity Lying Down.

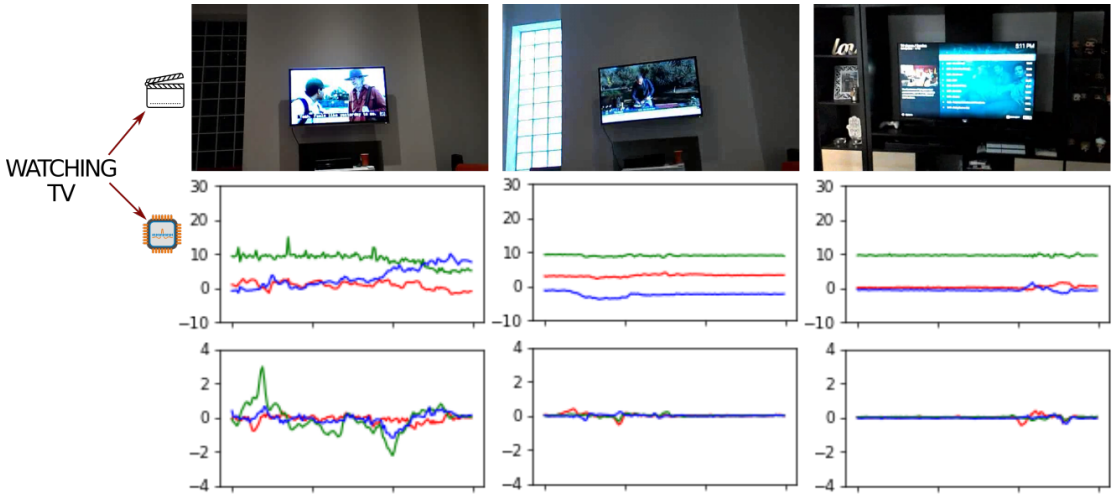

Figure 23: Watching TV.

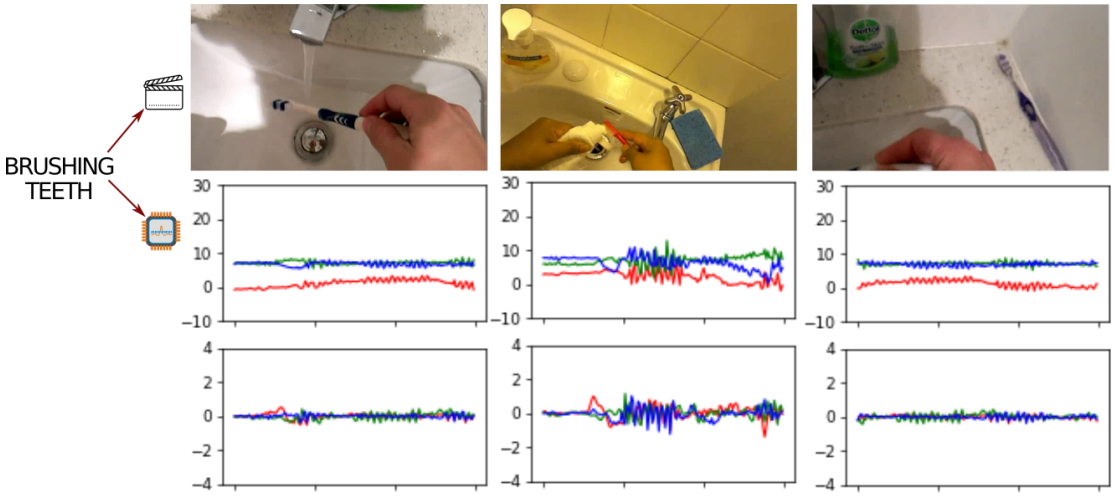

Figure 24: Activity Brushing Teeth.