D2.11 Dataset for composite human activities using a smartphone and a wrist-worn device

Project	SWELL
Project leader	Wessel Kraaij (TNO)
Work package	WP2
Deliverable number	2.11
Authors	Muhammad Shoaib
Reviewers	Saskia Van Dantzig, Edwin Matthijssen
Date	12/6/2015
Version	
Access Rights	
Status	

SWELL Partners:

Almende, Noldus, Novay, Philips, TNO, Radboud Universiteit Nijmegen, Roessingh Research and Development, Sense-Os, Universiteit Twente,

Summary

This document describes our dataset for composite human activities that can be used to detect good or bad habits in our daily life. Moreover, some of these activities can help coaching systems in providing feedback or coaching at the right time. This dataset is collected with 5 participants; who performed 13 different activities while carrying one smartphone in their pocket and another on the wrist, thereby emulating a wrist-worn device or a smart watch. Using our Android data collection app, data is recorded from four sensors of the Samsung Galaxy S2: an accelerometer, a linear acceleration sensor, a gyroscope and a magnetometer. The recognition of physical activities, smoking, eating, and drinking coffee can help in detecting bad or good habits in our daily life, whereas the recognition of working on computer, writing, and giving a talk can help in improving the feedback systems. We evaluated this dataset for recognizing composite human activities using various machine learning algorithms, which will be reported in D2.8. Moreover, this dataset can be used in future studies to build on top of our work and improve it in different ways. The goal of this deliverable is to present our raw dataset for public use by other researchers.

Contents

1 Composite Human activities Dataset

In this chapter, we describe our dataset for 13 different activities, which is an extended version of our previous dataset of 7 physical activities (deliverable D2.7). First, we describe how the data was collected. Then we present its structure. Finally, we describe how we used this dataset and its other potential uses.

1.1 Data collection protocol

In our data collection experiments, five users performed 13 different activities. All five participants were male and in their late twenties. Moreover, one of these participants was left handed. They carried smartphones (Samsung Galaxy S2 [1]) in their right jean's pocket and on their right wrist except the left-handed participant, emulating a smart watch. The left-handed participant used the smartphone in the right jean's pocket and the smartphone on the left hand. The body positions used in the data collection are shown in Fig. 1. Because smart-watches are typically equipped with sensors like an accelerometer and a gyroscope [2], we emulate a smart-watch using a smartphone (Samsung Galaxy S2) on the wrist position which is equipped with similar sensors.

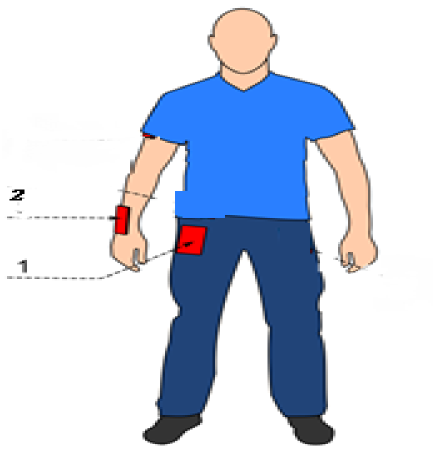

Fig.1. Smartphones in pocket and on wrist positions in data collection process.

Among the 13 activities, seven were simple physical activities, such as walking, jogging, biking, sitting, standing, walking upstairs, and walking downstairs. In sitting and standing activities, the user sat and stood still alone without talking and doing any other activity. The smartphones were used in same orientation (portrait orientation) on both positions. These simple seven physical activities were performed for 3 minutes by each participant, thereby producing 15 minutes of data for each activity. These users then performed 6 additional composite activities. The duration for these activities varied because different participants took different amount of time to complete them. . These activities are mentioned below with the amount of time (for all 5 participants) for which the data was recorded.

- ***Typing (21 minutes):*** *all five participants typed on their laptops the introduction section of one of our papers.*
- ***Writing (21 minutes):*** *they wrote the same introduction section on a paper of A4 size.*
- ***Drinking coffee (24 minutes):*** *they had a cup of coffee while sitting in office lounge alone.*
- ***Giving a talk (16 minutes):*** *they gave a talk in our meeting room about their research topic for 3-4 minutes.*

- ***Smoking (25 minutes):*** *Each participant smoked one cigarette while standing alone in the smoking area.*
- ***Eating (23 minutes):*** *For the eating activity, users were asked to eat soup or yogurt for 3-4 minutes in their natural style while sitting alone in the university cafe. The cup was placed on a table and participants used a spoon for eating.*

We collected data for multiple smartphone sensors, such as an accelerometer, linear acceleration, gyroscope, and magnetometer. A brief description of these is as follows [3,12]:

- ***Accelerometer:*** *it measures acceleration in meter per second squared (m/sec^2) along three axes.*
- ***Gyroscope:*** *it measures the angular velocity in radians per second (rad/sec) along three axes.*
- ***Magnetometer:*** *it measures the magnetic field in micro tesla (μT) along three axes.*
- ***Linear acceleration sensor:*** *it is obtained by removing the gravity component from the accelerometer values. It is also measured in meter per second squared (m/sec^2) along three axes.*

We collected the data at 50 samples per second. This sampling rate is enough to recognize human physical activities, as shown in our previous study [4], and in other relevant studies on activity recognition [5,6,7]. Moreover, lower frequencies have been shown to be sufficient for activity recognition [8, 9]. So this sampling rate can be down-sampled in future studies as per the needs of the other researchers. This dataset will be made available on our website [10] at a later stage. Our data collection app is already available on the website.

1.2 Dataset structure

There are 5 excel files in our dataset. Each excel file contains data for each participant's 13 physical activities for both positions. We present the dataset in its raw form so that it can be used in various ways in future studies by other researchers. This raw data can be preprocessed in different ways and various features can be extracted from it, depending on the goal of a study.

The following notations are used in these files as excel column headings besides timestamp:

- *For accelerometer: Ax = x-axis, Ay = y-axis, Az= Z-aixs*
- *For linear acceleration sensor: Lx = x-axis, Ly = y axis, Lz= Z-aixs*
- *For gyroscope: Gx = x-axis, Gy = y-axis, Gz= Z-aixs*
- *For magnetometer: Mx = x-axis, My = y-axis, Mz= Z-aixs*

1.3 Potential uses

We have used this dataset for recognizing composite human activities that can lead to detecting bad or good habits in our daily life. Some of these bad habits are the lack of physical activity, drinking too much coffee, smoking and missing meals. In this work, we extracted different time domain features from this data and used various machine learning algorithms to evaluate the recognition performance of these composite activities. This work has already been published and will be reported as deliverable D2.8. According to our evaluations, we found that simple physical activities can easily be recognized with one of these two positions. However, it is important to combine a smartphone with a wrist-worn device for recognizing composites activities in a reliable way, such as smoking while standing, eating while sitting and giving a talk. The fusion of motion data from these two positions gives us more context information and more reliability for the classification of all activities. We also found that the size of the window for feature extraction has an impact on the recognition performance of these activities. For simple physical activities, a smaller window of 2 sec was enough. However, for composites activities, relative bigger window size of 5 to 10 seconds was required for better activity recognition because 2 seconds was not enough to capture the repetitive gestures in these composite activities. This impact was also different for gyroscope and accelerometer. We also found that synchronization delays in sending data from wrist-worn device to smartphone do not affect the recognition performance in a negative way. Our results show that composite activities can be recognized in a reliable way with the combination of a smartphone and a wrist-worn device and such recognition can lead to detecting a higher level behavior such as good or bad habits. We have discussed all these different aspects of our evaluations in detail in [11], which readers can refer to for more details. However, this dataset can be used in different ways in future studies. Some of the examples are presented here [12]:

- *Investigating ways to improve the recognition performance of composite activities.*
- *Comparing various classification algorithms in different experimental setups.*
- *Evaluating the role of different smartphone sensors when they are used alone or in different combinations (sensor fusion).*
- *Comparing activity recognition performance using a smartwatch and a smartphone in pocket position.*
- *Analyzing position-independence, such that trained classifiers can be used on various body positions without training them on all these positions, thereby giving the user more freedom in wearing the devices.*
- *Comparing generic and personalized training and testing for classification algorithms. We provide each participant data separately so it can be used to train classifiers for individual users, thereby producing a personalized classification scenario.*

One of the limitations of this dataset is that it is not balanced in terms of the number of examples for each activity class in the training data. This is known as class imbalance problem [13]. The imbalance in activity classes can be seen in Section 1.1. For example, the total amount of data for smoking is for 25 minutes whereas that of giving a talk is for 16 minutes. Such class imbalance can lead to higher classification results for classes with more training examples because some classification algorithms are biased towards such majority classes [13]. Therefore, we are now extending this dataset for balanced class distribution, where each activity class will have the same number of training examples. We are currently evaluating this balanced dataset and will be made public on our website [10] in the future.

References

1. http://www.samsung.com/global/microsite/galaxys2/html/accessories.html [Last accessed on 6th June 2015]
2. http://www.lg.com/us/smart-watches/lg-W110-g-watch-r [Last accessed on 6th June 2015]
3. http://developer.android.com/guide/topics/sensors/sensors_overview.html [Last accessed on 6th June 2015]
4. Shoaib, Muhammad, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. "Fusion of smartphone motion sensors for physical activity recognition." Sensors 14, no. 6 (2014): 10146-10176.

5. Lara, Óscar D., and Miguel Labrador. "A mobile platform for real-time human activity recognition." In Consumer Communications and Networking Conference (CCNC), 2012 IEEE, pp. 667-671. IEEE, 2012.
6. V. Stewart, S. Ferguson, J. Peng, and K. Rafferty, "Practical automated activity recognition using standard smartphones, " in 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, 2012, no. March, pp. 229-234.
7. Incel, Ozlem Durmaz, Mustafa Kose, and Cem Ersoy. "A review and taxonomy of activity recognition on mobile phones." BioNanoScience 3, no. 2 (2013): 145-171.
8. Wu, Wanmin, Sanjoy Dasgupta, Ernesto E. Ramirez, Carlyn Peterson, and Gregory J. Norman. "Classification accuracies of physical activities using smartphone motion sensors." Journal of medical Internet research 14, no. 5 (2012): e130.
9. Anjum, Ashiq, and Muhammad Usman Ilyas. "Activity recognition using smartphone sensors." In Consumer Communications and Networking Conference (CCNC), 2013 IEEE, pp. 914-919. IEEE, 2013.
10. http://ps.ewi.utwente.nl/Datasets.php. [last accessed on 6th June 2015]
11. Shoaib, Muhammad, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. “Towards Detection of Bad Habits by Fusing Smartphone and Smartwatch Sensors”, PERCOM 2015 (PERCOM Workshops), ST. Louis, Missouri, USA.
12. Shoaib, Muhammad .“Extended Dataset for Physical Activity Recognition”. Deliverable D2.7, SWELL Project, 2014.
13. Japkowicz, Nathalie, and Shaju Stephen. "The class imbalance problem: A systematic study." *Intelligent data analysis* 6, no. 5 (2002): 429-449.