

# Adaptive Activity and Context Recognition using Multimodal Sensors in Smart Devices

Sébastien Faye, Raphael Frank, and Thomas Engel

Interdisciplinary Centre for Security, Reliability and Trust  
University of Luxembourg  
4 rue Alphonse Weicker, L-2721 Luxembourg  
{sebastien.faye,raphael.frank,thomas.engel}@uni.lu

**Abstract.** The continuous development of new technologies has led to the creation of a wide range of personal devices embedded with an ever increasing number of miniature sensors. With accelerometers and technologies such as Bluetooth and Wi-Fi, today's smartphones have the potential to monitor and record a complete history of their owners' movements as well as the context in which they occur. In this article, we focus on four complementary aspects related to the understanding of human behaviour. First, the use of smartwatches in combination with smartphones in order to detect different activities and associated physiological patterns. Next, the use of a scalable and energy-efficient data structure that can represent the detected signal shapes. Then, the use of a supervised classifier (i.e. Support Vector Machine) in parallel with a quantitative survey involving a dozen participants to achieve a deeper understanding of the influence of each collected metric and its use in detecting user activities and contexts. Finally, the use of novel representations to visualize the activities and social interactions of all the users, allowing the creation of quick and easy-to-understand comparisons. The tools used in this article are freely available online under a MIT licence.

**Key words:** Sensing System, Wearable Computing, Activity Detection

## 1 Introduction

In recent years, the growing availability and falling cost of smart devices have opened up a world of opportunities for new applications. Apart from smartphones, these connected objects include a wide range of ultra-portable devices that constantly interact with the users and their environment. Among those *wearables*, the vast majority are smartwatches and activity trackers. These have become very diverse and are equipped with high-performance sensors that allow users to monitor their physical activity in a way never possible before. Their sensors can read metrics from arm or hand movements with an accuracy comparable to specialized experimental devices [1]. These devices include physical sensors that are permanently in contact with the user's wrist, such as motion detectors (e.g. accelerometers) and environmental monitoring sensors (e.g. light sensors, microphone). Their ability to monitor other physiological metrics, such

as heart rate, leads to new areas of research. Further, the recent arrival on the market of major players, like Apple, Google and Microsoft, has facilitated the development and widespread adoption of sensing applications, opening the way to many new areas, including health, sport, and personal monitoring. According to ABI Research, it is estimated that the global market for wearables will reach 170 million in 2017 [2].

At present, whether we are talking about smartphones or wearables, these connected objects are generally used individually and for specific consumer applications (e.g. fitness). In most cases, the classic data fusion from sensors is adapted to be made in real time (e.g. pattern finding). This requires heavy-duty processing algorithms and consumes energy. Moreover, most systems only use smartphones, whereas wearables are more suitable for detecting user activities. Finally, few studies have looked at all types of existing sensors with the intention of arriving at a scalable and easy-to-implement solution.

In this paper, we intend to go one step further by presenting a sensing system that combines the data collected by one smartwatch and one smartphone. The platform that we have developed makes use of commercially-available devices and can be used to analyse the activity of a monitored user in great detail. Possible applications range from sports tracking systems to human behaviour modeling. Our contribution addresses four complementary objectives. (1) The design of an energy-efficient sensing system, using a streamlined fusion of data collected on two devices (a smartphone and a smartwatch). (2) The use of a supervised machine learning model to recognize user activities and their contexts. (3) The combination of multimodal metrics to obtain more advanced feature sets and generalize the recognition process. Finally, (4) the comparison of activities and social interactions of different users using novel 3D visual representations.

In the following section we provide a review of existing literature. Next, in Section 3, we present our sensing system, which is focused on the devices used for data collection and on how they communicate to exchange data. Section 4 describes our experimental campaign and how we used the collected metrics to create the data set used for our analysis. Section 5 focuses on the analysis of the data set and presents some relationships between metrics and a set of predetermined activities using a Support Vector Machine (SVM) model. These relationships form the basis for the recognition of activities and contexts to be inferred. Finally, two profile comparison methods are introduced in Section 6, before we conclude in Section 7.

## 2 Related Work

The use of mobile devices as key elements in a sensing system has been discussed for many years, both in industrial and research communities, as an opportunistic [3] or a participative system [4]. The classic architecture for such a sensing system consists of three parts [5, 6]. First, individual devices collect the sensor data. Then, information is extracted from the sensor data by applying learning

methods, generally on one of the devices or in the cloud, depending on the sensitivity of the data, the sampling strategy or the privacy level applied. Finally, the data can be shared and visualized from the cloud.

Smartwatches have their place in this kind of architecture and can open up new perspectives as they can collect the user’s activity and physiological signals [7], while smartphones are reserved for recording the user’s context. Smartwatches and smartphones are usually connected via Bluetooth Low Energy [8], a relatively new technology standardized under the Bluetooth 4.0 specification [9]. Compared to smartwatches, smartphones have a better battery capacity and can launch several tasks at the same time. By using a smartphone as a local gateway to access the Internet – via Wi-Fi or Cellular – we can turn this local sensing platform into a connected ecosystem [6].

As the applications need to be running on the devices permanently to collect and send data, there is an important compromise to be found between sample rate, rate of transmission and the consumption of energy [8]. The authors of [10] show, for example, that using all the sensors of a LG Nexus 4 E960 can reduce its battery from 214.3 hours (no sensors) to 10.6 hours (all sensors). Some systems attempt to circumvent this energy limit by offloading data processing onto servers [11]. Others propose sharing the data among neighboring phones [12]. By these means, cloud computing is widely used with smartphones and allows the creation of elastic models [13], where applications are launched on the mobile phone, and the data is processed in the cloud.

In the surveyed literature, accelerometers are the sensors most commonly used to recognize various physical and upper body activities. Indeed, [1] shows that specific movements of the arms, the hands and the fingers, generate sufficient energy to be distinguished by the accelerometer and the gyroscope in a smartwatch with 98% precision. By correlating different sources of data, other sensors such as GPS, microphones and Wi-Fi signals can also be used to improve the classification accuracy and estimate, for example, the mode of transport (e.g. bike, car) [14]. By continuously recording sound, it is possible to identify different user contexts, whether having a conversation, sitting in an office, walking out on the street or even making coffee [15, 16]. SPARK [17] is a framework that can detect symptoms associated with Parkinson’s disease using a smartwatch on the wrist (to detect dyskinesia using motion sensors), and a smartphone in the pocket (gait analysis and sound). Shin et al. [18] study patients with mental disorders and use smartwatches to help quantify the exercise and the amount of sunlight wearers have received, using GPS, accelerometer and the light sensor. Video sensing also permits various activities to be recognized [19]. However, video analysis is both algorithmically and computationally expensive, especially in a resource-constrained environment. Finally, social interactions can be identified using Bluetooth, Wi-Fi, Near-Field Communications (NFC) or cellular [10].

Activity detection involves the recognition of spatio-temporal patterns from sensor data that is usually incomplete and noisy. There is significant number of models that are able to characterize human behaviour from different features (e.g. accelerometer data). The temporal signal shape can be analyzed both in

time and frequency domains. Time-domain features include basic waveform characteristics and signal statistics that can be considered to be features of a given signal, e.g. the statistical moments, time between peaks, binned distribution, mean value of local maxima [20]. Data set reduction techniques such as Principal Component Analysis and Linear Discriminant Analysis can be used to extract the most significant discriminating features while reducing the dimensionality of the data representation [21]. Combining the feature extraction techniques above, activity recognition can be trained using (semi-)supervised methods in a controlled setting. These methods include Decision Trees, Neural Networks and Support Vector Machines, all of which have been successfully used in human activity recognition [22]. For example, Frame-based Descriptor and multi-class Support Vector Machine [23] is an approach that can classify a large variety of gestures. Unsupervised methods (e.g. k-means clustering [24]) can then be used to find structures in the different activity sequences and durations that were identified to find common properties or behaviours of user groups.

### 3 Sensing System

In order to carry out our studies and obtain the results presented in this article, we used our own system, SWIPE [7], which is available online<sup>1</sup> under a MIT licence. It is composed of two main parts: an Android application for data collection and a web platform for data processing.

#### 3.1 Hardware

We used two devices running Android 5.1.1. One was a smartwatch (Samsung Galaxy Gear Live) that records the wearer’s activity by registering wrist movements and other physiological data (i.e. heart rate). The other, a smartphone (LG Nexus 5), is responsible for collecting contextual data (e.g. with its microphone) as well as some additional activity data (e.g. accelerometer). The decision to run SWIPE on Android makes sense because of its maturity and its leading role in the current generation of smartwatches. Table 1 summarizes the specifications of the two devices, including details of the data that our system is able to collect.

#### 3.2 Architecture

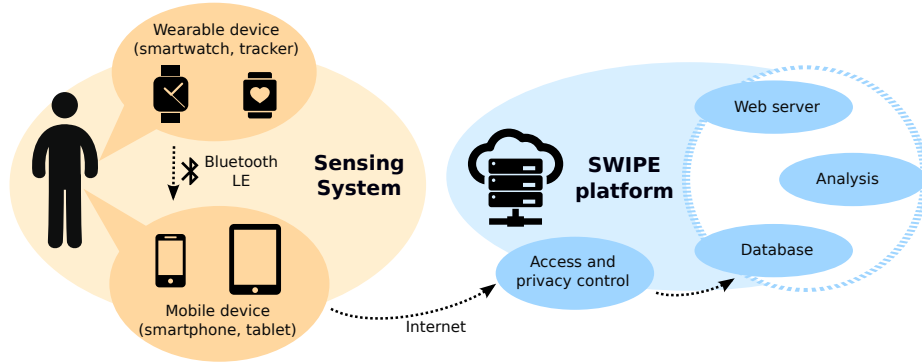
The architecture of SWIPE is shown in Figure 1 and consists of two parts. First, the sensing system is composed of a watch (worn on the wrist) and a phone (carried in the pocket) as introduced in 3.1. The watch periodically sends the data it has collected to the smartphone, which acts as a local collection point and as a gateway to access the SWIPE platform over the Internet (via Wi-Fi or a cellular network). The SWIPE platform is divided into several modules,

<sup>1</sup> <https://github.com/sfaye/SWIPE/>

Devices	RAM / Storage	CPU	Battery	Network Interfaces	Collectable Data
Samsung Gear Live	512 MB / 4 GB	Quad-core 1.2 GHz	300 mAh	Bluetooth 4.0	Heart rate, heart rate accuracy, pedometer, linear acceleration, time.
LG Nexus 5	2 GB / 16 GB	Quad-core 2.3 GHz	2,300 mAh	4G/LTE, GPS, 802.11 a/b/g/n/ac dual-band Wi-Fi, NFC, Bluetooth 4.0	Proximity, ambient light, linear acceleration, ambient sound, detected activity, activity confidence, pedometer, mobile network information, detected Bluetooth devices, Wi-Fi networks, GPS (location, altitude, speed, bearing, distance), local weather (from OpenWeatherMap.org), time.

**Table 1.** Specification of the devices used of our studies.

which (1) receive data following authentication and (2) store, (3) analyse and (4) display the data by means of a web interface. Each user is identified by a unique hash string and his or her data is stored on an internal University of Luxembourg server, which is accessible only on the local network. The link between the server and the sensing system is performed by an intermediate server that acts as a relay.



**Fig. 1.** SWIPE overall architecture.

### 3.3 Metrics Collected by SWIPE

The main metrics that our system collects are shown in Table 2. The “recording rate” column indicates the frequency at which a metric is saved in a data file, while the “sampling rate” indicates the frequency at which the system acquires raw data from sensors. Since the user is wearing the watch all the time, metrics associated with the watch include the ability to recognize activity. The average speed of movement of the user’s arm is recorded every 30 seconds, along with the maximum speed in order to detect sudden, unusual gestures. Metrics collected by the phone include contextual data. This includes accelerometer readings that are complementary to those provided by the watch. We also store microphone

readings to register the level of ambient noise, enabling us to distinguish between noisy and a quiet places. Network data also enables us to collect information on both mobility (GPS, Wi-Fi) and interaction with other users (Bluetooth).

Devices	Metrics	Sensors	Recording & Sampling rates		
Phone, watch	Maximum acceleration	Accelerometer	30 sec.	< 1 sec.	Maximum value of $\alpha = \sqrt{(x^2 + y^2 + z^2)}$ $m.s^{-2}$ , where x, y and z are the acceleration along each axis of the device, excluding gravity.
	Average acceleration	Accelerometer	30 sec.	< 1 sec.	Average value of $\alpha$ .
	Pedometer	Accelerometer, Android API	60 sec.	~	Number of steps taken by the user, detected by the Android system as a function of the accelerometer.
Watch	Heart rate	Optical heart rate sensor	60 to 300 sec.		Heart rate, in beats per minute, provided by the optical heart rate sensor.
Phone	Ambient sound	Microphone	60 sec.	1 sec.	Ambient sound level, from [0 : 100].
	Bluetooth devices	Network	120 sec.		Number of Bluetooth devices.
	Wi-Fi APs	Network	300 sec.		Number of Wi-Fi Access Points.
	Mobile network data state	Network	300 sec.		Value expressing the use of cell phone network. A reading of zero indicates that the phone is connected to a Wi-Fi AP. For our experiments, only the access point of our workplace was configured.
	Speed	GPS	60 sec.		Travel speed, in $km.h^{-1}$ .

**Table 2.** Key metrics collected by SWIPE.

### 3.4 Energy Saving Strategy

The provision of a sensing system launched as a background service represents a potential burden on the batteries of the devices used, which (particularly in the case of smartwatches) are not renowned for their longevity. It is therefore critical that we make every effort to save energy. This includes finding the right compromise between energy consumption and data collection. The proposed system aims to run uninterrupted for at least 12 hours in order to collect enough data to obtain an overview of daily activities. To achieve this, we implemented the following optimization strategy.

(1) Data transmission consumes a significant amount of energy. We first configure our application so that the watch, if close enough, uploads its data to the smartphone every 20 minutes rather than continuously. This allows the application to automatically turn off Bluetooth most of the time and makes the watch fully autonomous (i.e. the user can wear the watch without having to carry the phone). Data collected and transmitted by the smartwatch is received and stored locally by the smartphone to be sent once a day to our servers for later analysis. The data is sent at a predefined time (at midnight) or when the battery level of either of the devices drops below a threshold of 5%.

(2) Another factor that contributes to energy consumption is the frequency at which the sensors record data. The higher the frequency and the longer the transmission time, the more energy is consumed. On the other hand, a lower data acquisition rate will dilute the quality of the resulting data set. Consequently, each metric is configured with the parameters set out in section 3.3.

Note that while most of the metrics are configured with a fixed and adequate sampling frequency with respect to the tests carried out, other strategies are set up for specific cases. Indeed, the acquisition frequency of the heart rate sensor is designed to adapt to the activity of the user. When the user is making little or no movement, the sampling frequency is low, since his heart rate should be stable and the measurements reliable. Conversely, when the user moves, the sensor becomes more sensitive and his heart rate is likely to change. In this case, the data acquisition rate increases in order to take more probes.

(3) Finally, the devices are configured to prevent users from interacting with them. Each is locked with a password and all the unnecessary services managed by Android, such as notifications, are disabled. This allows us to record the data without interruption and under the same conditions for every participant.

This energy saving strategy is evaluated by comparing it with the settings where transmission, harvesting and recording frequencies were high (i.e. all set to 1 second). We find an autonomy gain of about 287% for the smartwatch (13.5h vs. 4.7h) and on the order of 189% for the smartphone (15.7h vs. 8.3h).

## 4 Building a Data Set

### 4.1 Scenario

The studies we conducted involved 13 participants working in the same building at the University of Luxembourg. These participants were selected as a representative sample of both genders and of different ages. Each participant was systematically subjected to the same requirements: (1) wear the watch and smartphone for one day, from 10:00 to 23:59; (2) complete a questionnaire<sup>2</sup> asking for an exact description of activities carried out (work, commute and leisure activities); (3) sign an informed consent form to accept the privacy policy of the study.

### 4.2 Example

Figure 2 shows data from one of the participants over a period of about 14 hours. The accelerometer data and the level of ambient noise immediately reveal several distinct situations. Around 19:00, for example, the participant appears to perform much faster movements than usual with both his watch and his phone – indicating that he is carrying both devices. The noise level is also high, indicating either a noisy place or excessive friction (which is common when the phone is carried in a pocket). We can easily deduce that the user was running. This is confirmed by the activity recognition algorithm provided by Android, which is able to detect basic activities. The situation is similar around 18:00. The environmental noise level is high, but both devices detect much less movement and the GPS records more rapid progress from place to place: the user was driving. These initial observations form the basis of our intuitive understanding of the user’s activity.

<sup>2</sup> Available online: <http://swipe.sfaye.com/mobicase15/questionnaire.pdf>

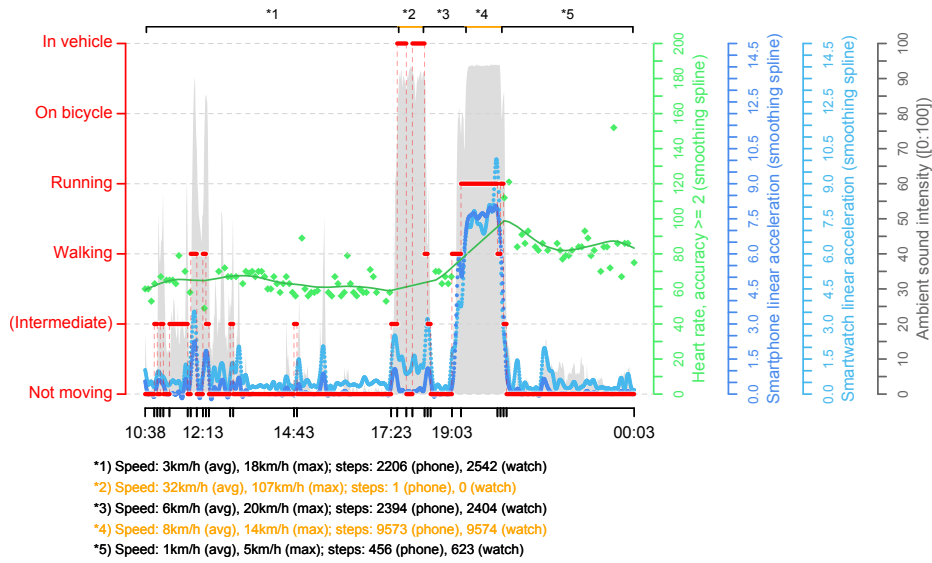


Fig. 2. Example of collected metrics for one participant.

### 4.3 Activity and Context Classes

In order to build a data set, we used both the information provided by users in the questionnaire and the information from the sensing platform. Each participant told us about the activities he or she had performed. By gathering all the information from the 13 participants, we obtained a total of nine activities (i.e. sitting, standing, walking, running, playing tennis, on a train, on a bus, on a motorcycle, in a car) that can be classified within five different contexts (i.e. working in an office, attending a meeting, in a shopping centre, on a break at work, at home), as represented in Table 3. Since we have the time slots for each activity (e.g. Figure 2), we are able to assign a set of representative values for each activity and context considering multiple inputs.

Activity									Context					Total
Physical					In vehicle				Working	Meeting	Shopping	Break	At home	
Sitting	Standing	Walking	Running	Tennis	Bus	Train	Motorcycle	Car	Working	Meeting	Shopping	Break	At home	Total
5.2h	11h	3.8h	1.9h	0.8h	1.4h	0.6h	0.3h	7.8h	5.2h	8.1h	3.4h	8.7h	7.3h	65.4h

Table 3. Identified activity and context classes with their total durations in our data set, which consists of 157.2h of recordings.



## 5 Activity and Context Recognition using SVM

### 5.1 Parameters

The problem to be solved is how to identify a class based on a set of metrics. We chose to use SVM (Support Vector Machine) [25], a set of supervised learning techniques to classify data into separate categories. SVMs have the ability to deal with large amounts of data while providing effective results. They can be used to solve problems of discrimination, i.e. deciding which class a sample is in, or regression, i.e. predicting the numerical value of a variable.

For our evaluation we used the SVM classifier provided by the `e1071` package for R [26]. The default optimisation method – C-classification – is used, as well as the classic radial kernel. Grid-search with 10-fold cross validation [27] was used to adjust the cost parameter C (within a range of 1 to 100), as well as  $\gamma$  (within a range of 0.00001 to 0.1).

### 5.2 Feature Set

The numerous measurements that we have in our data set were not all recorded at the same frequency. As shown in Table 2, acceleration was recorded twice as often as GPS speed. To simplify future operations, we chose to refine the data for each metric by sampling the same number of values from each. For each of the known classes selected in Section 4.3, we use a sliding window of ten minutes, moving over the data stream every five minutes. With each movement of the window, two representative values of data included in the window – referred to as  $x$  – are recorded: their average  $\bar{x}$ , which gives an overall view of the data over the interval; and their standard deviation  $\sigma(x)$ , which is fundamental to understanding the variations around the average. Finally, each activity and context class is represented as a set  $M$  of  $m$  metrics, each of which is represented, for each 10-minute data interval  $x$ , as  $\bar{x}$  and  $\sigma(x)$ . The following matrix illustrates the structure of the data set:

$$\begin{matrix} & \textit{Class} & M_1^{avg} & M_1^{std} & \dots & M_m^{avg} & M_m^{std} \\ T_1 & \left( \begin{matrix} class_1 & \bar{x}_{1,1} & \sigma(x_{1,1}) & \dots & \bar{x}_{1,m} & \sigma(x_{1,m}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ class_\Delta & \bar{x}_{n,1} & \sigma(x_{n,1}) & \dots & \bar{x}_{n,m} & \sigma(x_{n,m}) \end{matrix} \right) & & & & & \end{matrix} \quad (1)$$

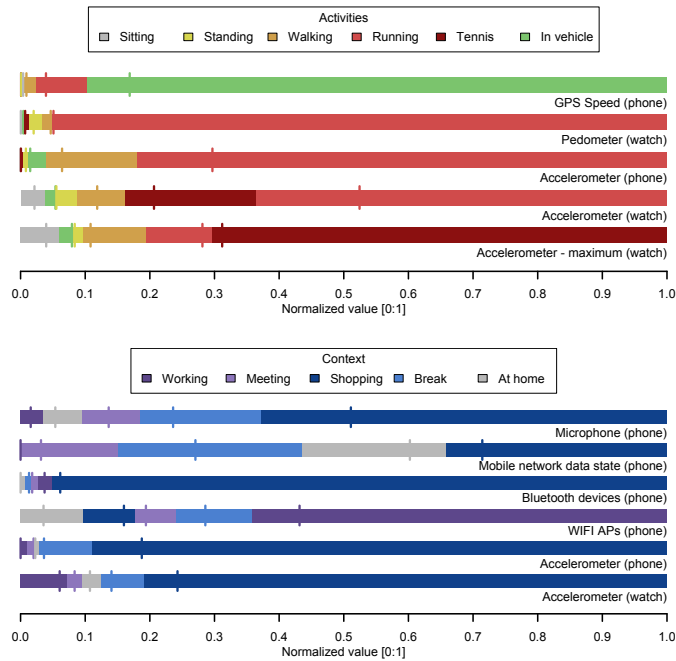
This representation is simple and has the advantage of abstracting from the excessive precision of the data. It also has the advantage of being lighter and less expensive to treat with a classification algorithm. Assuming we have a set of data composed of  $t$  seconds of recording, that the length of the sliding window is  $t_{window}$  seconds and that it moves every  $t_{step} \leq t_{window}$  seconds, we obtain a data matrix whose size is:

$$\textit{columns} = (2 \cdot m + 1) \quad \textit{rows} = \frac{t - (t_{window} - t_{step})}{t_{step}} \quad (2)$$

Our activities database contains, for example, a total of 65.4h of recordings and is  $19 \times 784$  in size.

### 5.3 Recognition using Metrics Individually

First of all, we investigate the individual influence that each metric can have on the recognition of an activity and/or context. Figure 3 represents some selected normalized metric averages over all participants and for each class. For reasons of visualization, the vehicle activities are grouped into the “In vehicle” class. The colour transition between each class represents half the distance that separates their average. The findings are logical, but they confirm the individual importance of each metric. For example, on average the GPS speed reading can help to detect whether the user is traveling in a vehicle, running or at rest. Maximum accelerometer readings can help us recognize a sport activity, such as tennis. Noise in a shopping centre seems to be higher than noise during a meeting.



**Fig. 3.** Selected metric averages for each class.

We want to use streamlined versions of the data set described in Section 5.2, with the aim of representing each metric individually to see whether or not it can accurately detect a class. Each data set is evaluated in order to discover how accurately we can predict a class based on a single metric. To do this, each data set is randomly divided into two parts. The first is the training set, comprising 70% of instances. The second is the test set, comprising the remaining 30%. The training set is subjected to a grid search to find the cost and  $\gamma$  that minimize the error rate. An SVM model is created from the training set using the best cost and the best  $\gamma$ . The model is then confronted with the test set with the

aim of predicting the number of instances in the test set whose class is correctly recognized by the training set. In order to ensure a representative average value for the error rate, this operation is performed 100 times for each combination and calculated for each iteration as  $1 - Accuracy$ . The results are shown in table 4 with  $Accuracy = \frac{true\ value}{total\ value}$ .

Device	Metric	Activity								Context					
		Physical				In vehicle				Working	Meeting	Shopping	Break	At home	
		Sitting	Standing	Walking	Running	Tennis	Bus	Train	Motorcycle						Car
Watch	Max. Acceleration														
	Acceleration														
	Pedometer														
	Heart rate														
Phone	Max. Acceleration														
	Acceleration														
	Pedometer														
	Microphone														
	Bluetooth devices														
	Wi-Fi APs														
	Mobile network data state														
	GPS speed														

Accuracy: 0% 100%

**Table 4.** Influence of each metric on the recognition of classes. The red to yellow gradient indicates high to low prediction accuracy. Grey indicates that no data was available for the performed activity.

We notice a huge disparity between all combinations of metrics and classes. The overall findings were quite polarized: some metrics can identify a class with very high reliability (e.g. the relationship between acceleration and running), while others cannot. Of course, the combinations shown are representative of our data set, where activities were taking place in an urban environment. For example, it is normal to see Wi-Fi sometimes taking particular prominence. This would probably not be the case if environments were more heterogeneous.

#### 5.4 Recognition using a Combination of Multiple Metrics

It makes sense to use a classifier such as an SVM when combining multiple metrics to deduce an activity, which can be seen as a more advanced feature set. We are interested in minimizing the error rate returned by an SVM model, that takes a set of metrics as its input, i.e. finding a combination that minimizes the error rate for both of the activity category and the context category. To do this, we generate all possible combinations of metrics and create a data set for each combination (e.g. watch acceleration and heart rate, Wi-Fi access points and GPS speed and Bluetooth devices, etc.). In the same manner and with the same parameters as above, for all possible combinations, each data set is randomly divided into a test set and a training set in order to calculate the average error

rate provided by the combination, over 100 iterations. The combination retained is the one with the minimum average error rate.

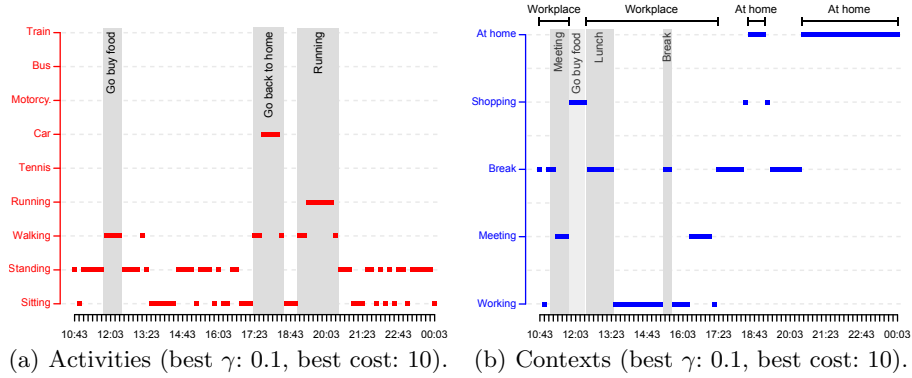
Category	Class	Best combination										Best accuracy					
		Watch					Phone					Phone & Watch	Phone only	Watch only			
		Max. Acceleration	Acceleration	Pedometer	Heart rate	Max. Acceleration	Acceleration	Pedometer	Microphone	Bluetooth devices	Wi-Fi APs				Mobile network data state	GPS speed	
Activity	Sitting														<b>99.5</b>	99.0	85.6
	Standing														<b>95.3</b>	81.4	68.8
	Walking														<b>95.8</b>	87.0	73.2
	Running														<b>100</b>	97.4	<b>100</b>
	Tennis														<b>100</b>	-	<b>100</b>
	Bus														<b>88.9</b>	69.5	73.1
	Train														<b>100</b>	96.6	83.9
	Motorcycle														<b>100</b>	97.8	<b>100</b>
	Car														<b>93.2</b>	90.4	77.0
	<b>Average</b>														<b>93.2</b>	88.1	81.5
Context	Working														<b>95.3</b>	<b>95.3</b>	72.3
	Meeting														<b>94.7</b>	85.8	73.4
	Shopping														<b>86.6</b>	76.7	80.6
	Break														<b>95.5</b>	84.9	69.4
	At home														<b>100</b>	99.5	92.4
	<b>Average</b>														<b>90.4</b>	86.6	74.7

**Table 5.** Best combinations of metrics for each activity. The “Best accuracy” columns denote the best possible percentage of the test data set which is correctly identified in the training data set. “Average” rows show the best combination for the entire class.

Table 5 represents the best combination of metrics obtained for each class of activity and context and for three cases: combined watch and phone metrics, watch metrics, and phone metrics. For each line, the best combination presented is the one that has the best accuracy. For example, the best combination for recognizing the “standing” class is a combination of metrics on the watch and on the smartphone, giving us a 95.3% average recognition accuracy. We can also see that, for the “running” and “motorcycle” classes, using the watch alone provides better accuracy than a combination of the watch and phone sensors. However, in most cases, the combined use of both devices offers better results than a phone or a watch alone. On the whole, the conclusions on the dataset are the same as those of 5.3. However, we can see that activity classes tend to be better served by motion metrics, whereas context classes are based more on Bluetooth, microphone, or network metrics. Finally, the two “Average” lines indicate a common combination in all classes that minimizes the average error rate. For example, the average context category combination is the one with the lowest average error rate for the classes of the category. These two lines are used in the next section to determine users’ classes.

## 5.5 Application Example

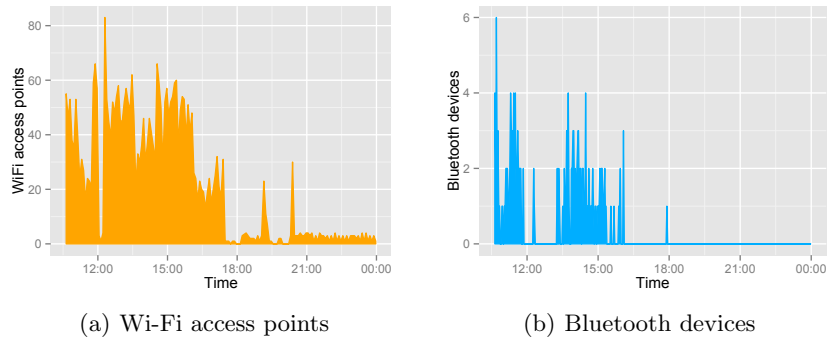
To illustrate our conclusions on the data set, we have taken as an example the participant shown in Figure 2. Each activity and context class is identified using the average combinations (Table 5). The recognition method is applied by progressively comparing the individual user’s data with the data in our full data set using SVM. Figures 4(a) and 4(b) illustrate the activity and user context recognition respectively, when the user’s data is not included in the full data set.



**Fig. 4.** Detected classes. Grey bars and black text are the main activities and contexts reported by the user.

The participant’s data is divided into ten-minute intervals. For each interval, we calculate the mean and the standard deviation of each metric. The set of values for the participant is small and therefore relatively easy to obtain. Each ten-minute interval consists of 14 values for activity detection and ten values for context recognition. As we can see from the figures and by consulting the participant’s questionnaire, we obtain a very realistic result, which is made possible by the collaboration of all participants and the pooling of their data. In Figure 4(a), for example, we see that at around 18:00 the participant was driving, and at between about 19:00 and 20:00 he was running. He took a lunch break around noon, which required him to move (walk) to buy food, as confirmed in Figure 4(b). The same figure also indicates that the participant was in his office most of the afternoon. Some errors are noted around 18:00, where it was detected that the participant was in a shopping centre, where in fact he was in his car. Similar findings are noted around 19:00, when the participant was running. The reason for this is the lack of a corresponding context class, and therefore the closest alternative is indicated.

Figures 5(a) and 5(b) respectively show the changes in the number of Wi-Fi access points and Bluetooth devices that the participant encounters. It is interesting to compare these figures with the previous ones, because they highlight certain geographic and social characteristics. In Figure 5(a), for example, there



**Fig. 5.** Geographic and social characteristics.

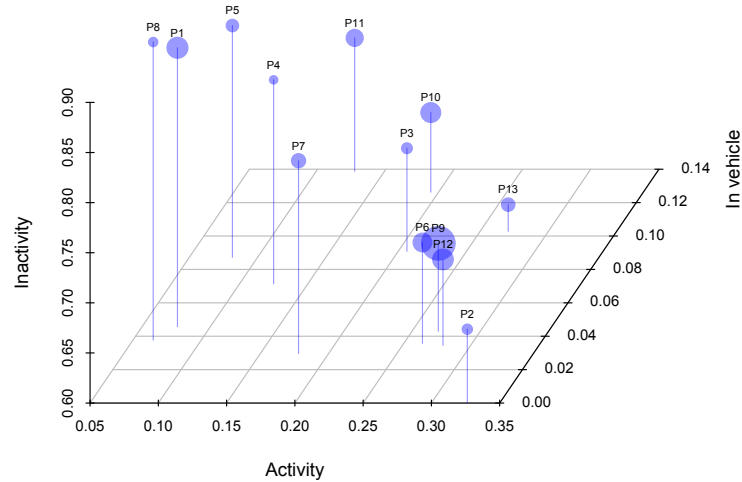
is a huge difference in the number of Wi-Fi access points encountered before and after 18:00, suggesting that the participant visited two major places (in this case, a work environment and a domestic one). It is also interesting to observe the dip around 12:00, which is when the participant visited the shopping centre. The participant’s movement, by car around 18:00 and while running between around 19:00 and 20:00, is also indicated by some slight spikes that we are often associated with travel: the more a participant moves, the more he comes into proximity with different access points. However, these figures do not provide a particularly accurate information base for estimating the participant’s social interactions. To do this, in the following section, we compare activities and social interactions among the participants.

## 6 Comparing Participants

If the recognition of user activity is an essential step that we can approach with great accuracy, another critical step is to compare several participants. In this section, we introduce novel visual representations, allowing comparison of the 13 participants in the study.

Figure 6 is a 3D plot showing the distribution of types of activity following three different axes. The first reflects the proportion of time the participants were inactive (e.g. sitting). Because the measurements were taken during workdays, the proportion is very high and goes from 63% (P13) to 90% (P8). The second reflects the proportion of time the participants were active and were performing an activity (e.g. walking, running). This number distinguishes two categories: those with a sporting activity outside of work and those who are required to move (e.g. to meetings). The third axis reflects the proportion of time the participants were aboard a vehicle. This number is the lowest, and corresponds mainly to journeys between work and home. However, participants such as P10 or P11 have work activities involving frequent trips during the day (e.g. to move from one campus to another). Finally, note that the size of a bubble is proportional to

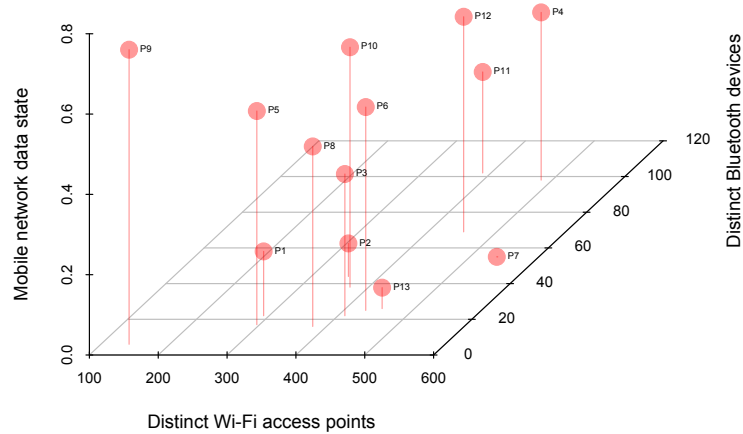
the sum of all acceleration recorded by the watch. Thus, a small bubble indicates very little sports activity while a larger bubble indicates more frequent, abrupt movement (e.g. running).



**Fig. 6.** Comparing activities of the participants.

Figure 7 uses the same principle as the previous figure but is based on three network metrics. First, the average mobile network data state tends to 0 if the mobile phone is connected to a Wi-Fi access point and it tends to 1 if the mobile phone uses cellular data. As the devices are set up only to connect to workplace access points (Table 2), this value is a good indicator of whether the user is more likely to be in the workplace or outside. The number of different access points gives us information about geographic locations visited by the participants. If two people are working in the same place, the participant with the higher value is moving around more and coming into contact with more access points. Finally, the number of distinct devices encountered gives us a measure of the interaction that the participants have. The higher this number, the more devices (a proxy for people) the person has encountered during his or her recording session.

Comparing the two graphs allows us to make some interesting observations. For example, participant P9 seems to perform more physical activity than anyone else, judging from his relatively high activity rate. Moreover, looking at figure 7, we find that P9 does not spend much time at the workplace, as he or she encounters the lowest number of access points. Conversely, participant P7 was mainly working during the study and hardly moved at all. Participant P4 is an interesting case, since he or she seems to have been in a vehicle and been in the proximity of a large number of access points. This indicates movement through many public spaces or buildings.



**Fig. 7.** Comparing interactions of the participants.

## 7 Conclusion

In this paper, we have described a strategy for recognizing the activities and the contexts within which a user is located. Our results show that using a condensed data set, along with energy-efficient sampling parameters, has the advantage of being easy to use with a classification algorithm such as SVM. Moreover, as such a structure implies lower transmission, harvesting and recording frequencies, it allows energy savings (resulting in an autonomy of about one day using our sensing system). We then showed that using a smartwatch in addition to traditional smartphones leads to better detection accuracy, in particular regarding physical activities such as running (100% accuracy over our dataset) or walking (95.8%). In addition, as these wearables are permanently on the user’s wrist, they can detect specific activities without the help of any smartphone (e.g. tennis). Overall, the use of multimodal metrics as advanced feature sets for an SVM model allows the recognition of nine user-defined activities and five contexts, with an average accuracy greater than 90%. Finally, we presented a new approach that graphically compares the activity and social relations of different users, allowing a better understanding of their behaviour.

The relatively small number of participants and their often vague answers to the questionnaire prevented us from expanding our data set. However, the study suggests great potential for the detection of personal activities if carried out on a wider sample group of users. In future work, in addition to using new devices and extending our energy saving strategy, we plan to carry out similar tests on a larger scale, performing new experiments and/or using public data sets. This will not only allow us to use other learner types and refine our classification model (e.g. adding FFT-based features), but also to accumulate a more extensive activity database that can be used as training set. We also plan to extend our study to capture user activities and contexts on a weekly basis, which would further help us to recognize patterns and characteristics specific to each user.



## References

- [1] C. Xu, P. H. Pathak, and P. Mohapatra. “Finger-writing with Smartwatch: A Case for Finger and Hand Gesture Recognition using Smartwatch”. In: *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. ACM. 2015, pp. 9–14.
- [2] S. Tilenius. *Will An App A Day Keep The Doctor Away? The Coming Health Revolution*. *Forbes CIO Network*. 2013.
- [3] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson. “People-centric urban sensing”. In: *Proceedings of the 2nd annual international workshop on Wireless internet*. ACM. 2006, p. 18.
- [4] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. “Participatory sensing”. In: *Center for Embedded Network Sensing (2006)*.
- [5] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. “A survey of mobile phone sensing”. In: *Communications Magazine, IEEE* 48.9 (2010), pp. 140–150.
- [6] F. Yang, S. Wang, S. Li, G. Pan, and R. Huang. “MagicWatch: interacting & segueing”. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM. 2014, pp. 315–318.
- [7] S. Faye and R. Frank. “Demo: Using Wearables to Learn from Human Dynamics”. In: *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM. 2015, pp. 445–445.
- [8] X. Zheng and J. Ordieres-Meré. “Development of a human movement monitoring system based on wearable devices”. In: *The 2014 International Conference on Electronics, Signal Processing and Communication Systems (ESPSCO 2014)*. 2014.
- [9] S. Bluetooth. “Bluetooth specification version 4.0”. In: *Bluetooth SIG* (2010). URL: <http://www.bluetooth.org/en-us/specification/adopted-specifications>.
- [10] J. G. Rodrigues, A. Aguiar, and J. Barros. “SenseMyCity: Crowdsourcing an Urban Sensor”. In: *arXiv preprint arXiv:1412.2070* (2014).
- [11] E. Cuervo, A. Balasubramanian, D.-k. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl. “MAUI: making smartphones last longer with code offload”. In: *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*. ACM. 2010, pp. 49–62.
- [12] R. Honicky, E. A. Brewer, E. Paulos, and R. White. “N-smarts: networked suite of mobile atmospheric real-time sensors”. In: *Proceedings of the second ACM SIGCOMM workshop on networked systems for developing regions*. ACM. 2008, pp. 25–30.
- [13] S. Hussain, J. H. Bang, M. Han, M. I. Ahmed, M. B. Amin, S. Lee, C. Nugent, S. McClean, B. Scotney, and G. Parr. “Behavior Life Style Analysis for Mobile Sensory Data in Cloud Computing through MapReduce”. In: *Sensors* 14.11 (2014), pp. 22001–22020.

- [14] M. Han, Y.-K. Lee, S. Lee, et al. “Comprehensive context recognizer based on multimodal sensors in a smartphone”. In: *Sensors* 12.9 (2012), pp. 12588–12605.
- [15] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell. “Sound-Sense: scalable sound sensing for people-centric applications on mobile phones”. In: *Proceedings of the 7th International Conference on Mobile systems, Applications, and Services*. ACM. 2009, pp. 165–178.
- [16] L. Ma, D. Smith, and B. Milner. “Environmental noise classification for context-aware applications”. In: *Database and Expert Systems Applications*. Springer. 2003, pp. 360–370.
- [17] V. Sharma, K. Mankodiya, F. De La Torre, A. Zhang, N. Ryan, T. G. Ton, R. Gandhi, and S. Jain. “SPARK: Personalized Parkinson Disease Interventions through Synergy between a Smartphone and a Smartwatch”. In: *Design, User Experience, and Usability. User Experience Design for Everyday Life Applications and Services*. Springer, 2014, pp. 103–114.
- [18] D. Shin, D. Shin, and D. Shin. “Ubiquitous Health Management System with Watch-Type Monitoring Device for Dementia Patients”. In: *Journal of Applied Mathematics* 2014 (2014).
- [19] L. Porzi, S. Messelodi, C. M. Modena, and E. Ricci. “A smart watch-based gesture recognition system for assisting people with visual impairments”. In: *Proceedings of the 3rd ACM international workshop on Interactive multimedia on mobile & portable devices*. ACM. 2013, pp. 19–24.
- [20] Z. He, Z. Liu, L. Jin, L.-X. Zhen, and J.-C. Huang. “Weightlessness feature—a novel feature for single tri-axial accelerometer based activity recognition”. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE. 2008, pp. 1–4.
- [21] T.-P. Kao, C.-W. Lin, and J.-S. Wang. “Development of a portable activity detector for daily activity recognition”. In: *Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on*. IEEE. 2009, pp. 115–120.
- [22] H. Qian, Y. Mao, W. Xiang, and Z. Wang. “Recognition of human activities using SVM multi-class classifier”. In: *Pattern Recognition Letters* 31.2 (2010), pp. 100–111.
- [23] J. Wu, G. Pan, D. Zhang, G. Qi, and S. Li. “Gesture recognition with a 3-D accelerometer”. In: *Ubiquitous intelligence and computing*. Springer, 2009, pp. 25–38.
- [24] A. K. Jain. “Data clustering: 50 years beyond K-means”. In: *Pattern recognition letters* 31.8 (2010), pp. 651–666.
- [25] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [26] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, and A. Weingessel. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*. 2011. URL: <http://CRAN.R-project.org/package=e1071>.
- [27] J. Bergstra and Y. Bengio. “Random search for hyper-parameter optimization”. In: *Journal of Machine Learning Research* 13.1 (2012), pp. 281–305.