Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 1135–1146

DOI: 10.15439/2015F359
ACSIS, Vol. 5

A generic framework to support participatory surveillance through crowdsensing

Apostolos Malatras, Laurent Beslay
European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen
Email: apostolos.malatras@jrc.ec.europa.eu, laurent.beslay@jrc.ec.europa.eu

***Abstract*—Harnessing the power and popularity of participatory or opportunistic sensing for the purpose of providing added value security and surveillance services is a promising research direction. However, challenges such as increased privacy concerns, as well as technological issues related to the reliable processing and meaningful analysis of the collected data, hinder the widespread deployment of participatory surveillance applications. We present here our work on addressing some of the aforementioned concerns through our related participatory application that focuses on crisis management and in particular buildings' evacuation. We discuss the technical aspects of our work, the viability and practicality of which is validated by means of a real experiment comprising 14 users in the context of an emergency evacuation exercise.**

I. Introduction

RECENT developments regarding the capabilities of smartphones that are increasingly equipped with middle- to high-end sensors and their widespread penetration in modern society have spawn a novel paradigm of information generation and sharing, that of participatory sensing [1]. In this bottom-up paradigm illustrated in Figure 1, users of smartphones take advantage of the capabilities of the devices that they are carrying in terms of sensing and collect data regarding their surrounding environment and themselves, e.g. acceleration, temperature, light, sound, etc. They then proceed with sharing this information with other users either by uploading it to a common repository accessible to everyone (perhaps in the form of a map service where the location of the collected data is also pinpointed), or by sending their data to a centralized entity that provides them with related services [2].

The paradigm of participatory sensing is applicable to a wide range of application domains. Of particular interest is its consideration in light of security applications, in which case a new research domain, i.e. that of participatory surveillance, emerges [3]. Participatory surveillance refers to the use of principles from participatory sensing in order to monitor, control, and assess a variety of events for the purpose of security [4]. For example, the evacuation of a building could be enhanced when having access to data collected from the evacuees through their smartphones or in case of criminal activities the legal and police authorities could have a wealth of data coming from smartphones of nearby people to support their investigations [5], even in the absence of operational networking infrastructures. The latter data yield no significant information as such, but subject to processing using machine learning techniques they could be used to deduce useful knowledge about the activities the users were conducting at the time of data collection.

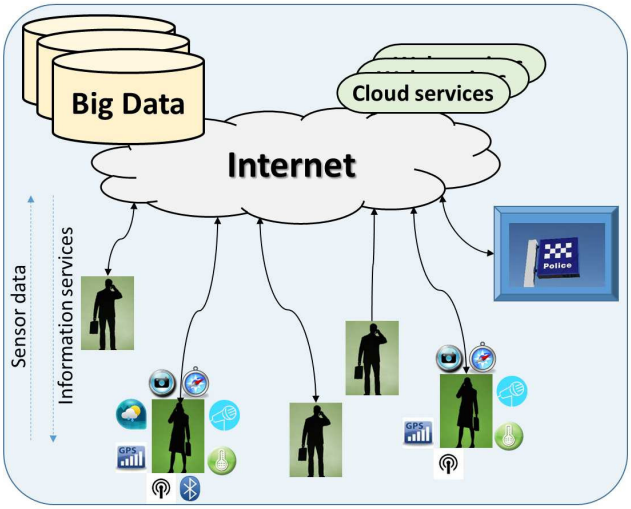

Fig. 1. Principles of operation of participatory surveillance.

To examine the viability of such scenarios we built a prototype participatory surveillance application and staged an evacuation exercise to validate its viability and practicality in real settings. We also designed a participatory surveillance framework and an experimental methodology to collect and analyze the data gathered throughout the exercise. One of the main goal of this research work was to examine the potential knowledge that can be derived from raw data stemming from the participatory sensing tasks. The driving objective was our ability to infer the different types of activities that the users were engaged in during the experiments, using only raw sensor data as input. We were also interested in examining how much information can be extracted from a minimal set of data and the results discussed later were quite interesting in terms of privacy.

We report here on our findings utilizing artificial intelligence and in particular machine learning algorithms to preprocess and analyze the collected data in order to infer the type of activities the users were conducting when the data was being collected. The results are very promising, with various configurations of our framework being able to identify up to

978-83-60810-66-8/$25.00

99% of users' activities (walking, standing still, climbing stairs or descending stairs) on the actual data collected from the evacuation exercise.

The remaining of this paper is structured as follows. After this brief introduction, Section II reviews related work in the area of participatory sensing in the context of security. Section III discusses the design of our participatory surveillance application together with potential scenarios of its use, whereas Section IV outlines privacy concerns and considerations regarding participatory surveillance. Section V presents the proposed generic framework to support analysis of data collected through such applications, evaluation of the accuracy and performance of which is the subject of Section VI. The paper concludes with Section VII where the limitations of this work are underlined and opportunities for further research in the domain are pinpointed.

II. Related Work

Sensors on mobile phones can be used to infer different types of information regarding the users of the phones, as well as the surrounding environment. Accordingly, an extensive review is presented in [6]. Accelerometers, gyroscopes, and other sensors have been used to detect human activities with typical cases presented in [7], [8], [9]. Such works are significant for security reasons, because they can reveal the state of users, e.g. running or laying still, and thus in combination with other contextual information they can hint on possibly suspicious actions of users, e.g. running away from a crime scene, while everybody else around the user is walking. Moreover, considering a smartphone's microphone as input, audio recordings regarding noteworthy events could be recorded in an inconspicuous manner, whereas the phone camera could serve as an additional channel of information reporting, as discussed in [10].

The prominence and ubiquity of current smartphones that are equipped with a variety of embedded sensors has spurred applications related to user-centric sensing and monitoring of the surroundings, namely the paradigm of participatory sensing [1]. Such applications have found applicability in environmental monitoring [11], green vehicle routing [12], noise mapping in urban environments [13] and lately in security and surveillance operations [4]. Despite the fact that there are several potential shortcomings from such an approach, i.e. regarding privacy [14] and the quality and accuracy of data collection [15], its benefits nonetheless are far from negligible.

In the context of security, participatory sensing applications can provide great data collection services at high granularity (spatial and temporal) and at a low cost. Current surveillance practices rely mostly on video monitoring, e.g. CCTV, which has several shortcomings in person identification [16] and cannot inherently cover extended areas. While there has been work on using sensors as information side channels for security operations, e.g. microphone [17], accelerometers and PIR [18], and magnetometers [19], such techniques have however not been envisaged at a large scale. Limited number of sensors were deployed in previous works and such approaches therefore also suffer from poor range and poor data records. Participatory sensing builds on the use of sensors on smartphones that are nowadays pervasive and ubiquitous and can thus provide information for large geographical areas, assuming a large volunteer user base. The associated costs are minimal compared to the installation and deployment of infrastructure-based solutions, while additionally the ease of deployment is great since nowadays users carry their phones with them for the major party of the day and it is always on, collecting sensor data. Another benefit of participatory sensing applications for security purposes can be found in the straightforward identification of people that is supported through the cell IDs of the phones and allows the association of users with the data related to them. The *Cell-All* project from the US Department of Homeland Security on the use of chemical sensors on mobile phones to detect chemical attack related emergency situations was one of the first efforts towards crowdsensing being put in use for security purposes [20].

Furthermore, one of the biggest concerns in participatory sensing is ensuring that users are actively contributing and sharing their data [21], [22] since it could eventually lead to poor performance due to the lack of accurate and informative representations. This problem has been considered in the context of noise mapping, where a persuasive, motivating game was considered in [23] to stimulate user data collection and sharing. In terms of participatory surveillance, we postulate that the citizens' sense of engagement and contribution in securing their environments will be the driving factor for their engagement. Nevertheless, in order to promote such engagement the fundamental issue raised by users, i.e. privacy, should be addressed. In this respect, there has been significant research work on the anonymization of shared user data with prominent examples being reported in [24] and [25].

III. Participatory Surveillance Scenarios

Participatory surveillance as a concept aims at utilizing the notions of participatory sensing and the ubiquity of smartphones equipped with a wealth of sensors in order to provide services related to surveillance and public security. This paradigm shift aims at empowering both the citizen and the police authorities and raising public awareness and citizen engagement [3]. Citizens on one hand feel more empowered since they are contributing in securing their neighborhoods and acquire thus a more active role in their society. Police authorities on the other hand gain from the wealth of data coming from citizens' smartphones and other monitoring means. It is an inexpensive and efficient way of enriching the data coming from traditional sources of surveillance, e.g. CCTV, as well as reaching areas where deployment of CCTV-like systems is not possible or allowed. Moreover, data coming from smartphones is not only limited to pictures or videos, but it can also include data from the embedded sensors such as accelerometer, gyroscope, pressure, magnetometer, etc. This kind of data supports the police authorities in gaining a better understanding of potentially noteworthy incidents [4].

In this respect, we considered a scenario that involved a variety of actors in order to collect experimental datasets from smartphone sensors and also to test the feasibility of the notion of participatory surveillance. We chose to avoid scenarios that would seem invasive and that might be considered as threatening citizens' privacy, e.g. continuous localization using GSM signals or recording the audio and video signals surrounding citizens at all times. The scenario we considered involved crisis management and in particular an evacuation exercise of an office building in case an emergency occurs, e.g. fire. People inside the building, namely employees, safety staff, building delegates (in charge of enforcing standard evacuation procedures) and visitors, were assumed to have a smartphone equipped with a custom application that monitored the values registered from their sensors and reported the data back to a centralized control room.

In terms of participatory surveillance, the principal goal was the utilization of the collected sensor datasets to provide the remotely located control room supervisors with useful knowledge regarding the progress of the evacuation, e.g. bottlenecks in exits indicated by users standing still, running in stairs indicating panic, users in peculiar situations (lying down or falling). Accordingly, the collected datasets that contained raw sensor data, e.g. from accelerometers or gyroscopes, had to be processed and analyzed in a manner that would allow us to extract useful information about user activities related to the collected data. The motivation behind this exercise was to examine whether data from smartphones alone would be sufficient to support surveillance tasks during an emergency when infrastructure surveillance mechanisms such as CCTV would be unavailable, namely to study the potential use of smartphones and participatory applications as a backup channel for surveillance.

The evacuation exercise scenario is in our view typical of prospective participatory surveillance applications, since it exhibits the majority of desired characteristics. In particular, it considers the use of a variety of sensors and a large number of people; it is privacy-friendly since it allows people to decide when and what type of data they wish to share; it is easily extensible to include further features and data sources; it provides solid motivation for the use of smartphones for security operations, since in such a case the lack of infrastructure surveillance network would be detrimental to police operations. Undoubtedly, the major concern of user privacy is present in this scenario as well as in every other participatory sensing application, albeit at a smaller scale. In the following, we discuss relevant privacy concerns and describe our approach in alleviating them.

IV. Privacy Considerations

While the benefits stemming from participatory surveillance applications are evident, the privacy risks involved are not clear and need to be carefully considered. The mere concept of participatory surveillance comes along with a series of potential privacy risks. Users are required to share personal data coming from the sensors embedded on their phones, in order to support and improve security operations and promote the communal sense of safety. We illustrate in Section VI that even with the use of data coming from just the accelerometer, it is possible to infer the activities that the user was conducting at the time of data collection. Use of additional sensors could exacerbate this risk, providing more detailed information on user activities. Indicative of relevant privacy risks, is the recent work presented in [26] that considered RF-sensing to infer the state of device-free individuals without their cooperation. Taking into account the capabilities of modern smartphones and the wealth of data available to participatory surveillance applications, it becomes clear that proper privacy enhancing technologies need to be put in place.

A major concern refers to the fact that user data can easily be traced back to their owner, because of the nature of cellular networks and the uniqueness of phone identifiers. Since data can be used to expose potential private user information, anonymization techniques need to be utilized to hinder such exposure. A comprehensive review of related solutions can be found in [24], with the most prominent approaches considering techniques such as k-anonymity and l-diversity [25]. Moreover, the entire space of sensors on smartphones needs to be carefully examined. The latter are nowadays carrying a large number of sensors and the knowledge that can be extracted from them (by processing the corresponding sensed data) is still not fully chartered. Studies like the one presented in this paper, expose the privacy risks related to the accelerometer, however the need to perform similar studies for the entirety of available sensors is paramount. Accordingly, guidelines could be provided to the end users to instruct them on the potential risks involved in sharing data coming from diverse sensors. This is particularly important since users are quite sensitive about sharing photos or location data for example, but are unaware of the risks involved in sharing data from low level sensors that could be equally detrimental to their privacy [14].

Access to this data should also be protected, so as not to allow its unauthorized viewing and processing. Participatory surveillance is the context in which this data is being collected and therefore police authorities should access this data under this particular context. The role of national Data Protection Authorities (DPAs) naturally emerges as a safepoint of supervisory control regarding participatory data protection. In addition to any such effort, participatory surveillance applications should also be designed with privacy in mind, in which case they should record any access to data and the processing method invoked on them to assist in prospective inquiries. Furthermore, who will have ownership of this data and for how long it should be kept and processed remains an open issue that could trigger conflicting situations. One could for example postulate that data should be retained by police authorities only when a criminal activity took place and then stored indefinitely. However, the exposure of such an activity cannot be foreseen and therefore data should be kept to ensure its availability if needed. Another conflict that might arise involves the user wishing to remove his shared data (right to be forgotten), whereas the police authorities might

not permit this due to ongoing or forthcoming investigations. There is no panacea to resolve such conflicts, with application- and context-dependent solutions usually being the norm. Undoubtedly, appropriate legislation and rule systems should be introduced to regulate this newly established field and thus promote its prosperity.

V. Data Analysis Framework Design

The notion of participatory surveillance refers to the collection of data from a variety of smartphones and other mobile devices and in particular from the on-board sensors carried by such devices. Processing and analyzing this wealth of data could lead to the inference of interesting information and knowledge regarding various surveillance aspects, namely the identification of distinct human activities, the location of a user and his/her surroundings (physical and social) and the occurrence of abnormal conditions, e.g. extreme sound levels potentially attributed to screaming or intense physical stress possibly attribute to user falling. The goal of the data analysis part of any effective participatory surveillance system is therefore to deduce such useful information. In the following we present a generic framework that has been devised for such purposes of data analysis and discuss its various aspects.

Analyzing data to extract useful patterns and accordingly use these to identify human activities and distinguish between them has been a very active research domain over the years [27]. The goal is to have computing systems capable of inferring knowledge by themselves using only raw data as input. The main elements of machine learning for the purposes of activity recognition using smartphones include the following:

- Data collection: collection of data using sensors located on the users' smartphone.
- Data training: the collected raw data need to be processed in order to deduce some useful information features and characteristics that will assist in its classification.
- Data classification: use of the aforementioned features in conjunction with machine learning classification algorithms to classify data, i.e. assign classes to data instances.

Figure 2 depicts these 3 different steps in the machine learning process. We applied these steps on both the collected reference data and the test data for our experiments and appropriately configured the classification process to enhance its performance in respect to our requirements. During the training phase the most appropriate and appropriately configured classifier is selected, so as to be applied in the testing phase over new data and classify them accordingly.

As previously discussed, sensor data can be exploited to detect human activities and thus provide insight on the actions and whereabouts of the smartphone users in a non-intrusive manner. We decided to use a systematic approach to tackle this problem and for this reason introduced a generic framework for the analysis of data coming from participatory surveillance activities. The main goal is to build a comprehensive dataset for training statistical classifier and applying this to actual, i.e. test, data to establish possible patterns/matches and thus identify human activities for the purposes of surveillance. The reason why we chose to define a generic framework is to establish a methodological framework of doing similar experiments, as well as to facilitate further developments in the domain. The framework is presented in Figure 3, while its elements are detailed in the following:

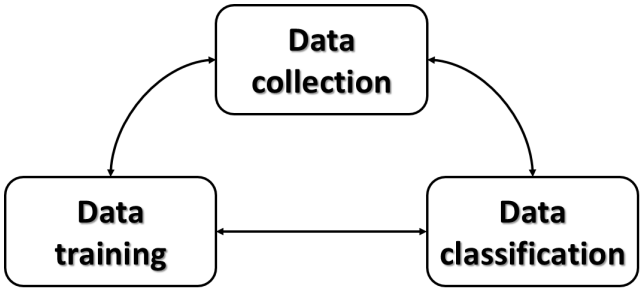

Fig. 2. Machine learning stages for activity recognition applied on both training and testing phases.

- Define problem space: the particulars of the participatory surveillance tasks need to be carefully defined at a high level, namely what needs to be achieved. They will serve as the requirements that will drive the rest of the analysis process.
- Define activities to be identified: not all scenarios for participatory surveillance rely on the identification of the same set of activities. However, the selection of the interesting activities is important at an early stage since it drives the definition of the required data and sensors to monitor these activities.
- Define sensors to be used: having defined the activities we are interested in, the next step concerns the selection of the most appropriate sensors to support the identification of these activities.
- Plan and conduct training experiments: the training phase is the first phase in the machine learning process and it involves a set of base experiments to collect reference data for the activities in question. The planning of these experiments is therefore of paramount importance, so as no configuration parameters or testing conditions become neglected during the following phase.
- Collect datasets for training: collect training data referring to elementary activities related to participatory surveillance as defined in previous steps of the process. The data should refer to more than one repetitions of the activity over a span of time.
- Pre-process training data: the collected data is in most cases noisy and needs to be pre-processed prior to being used by machine learning classifiers. Raw data usually has very fine granularity making it difficult to discern relevant statistical properties. Therefore it needs to be processed in order to extract statistical features over time, as well as in the frequency domain. The pre-processing tasks involve the removal of outliers from the original dataset, the annotation of the data for classification pur-

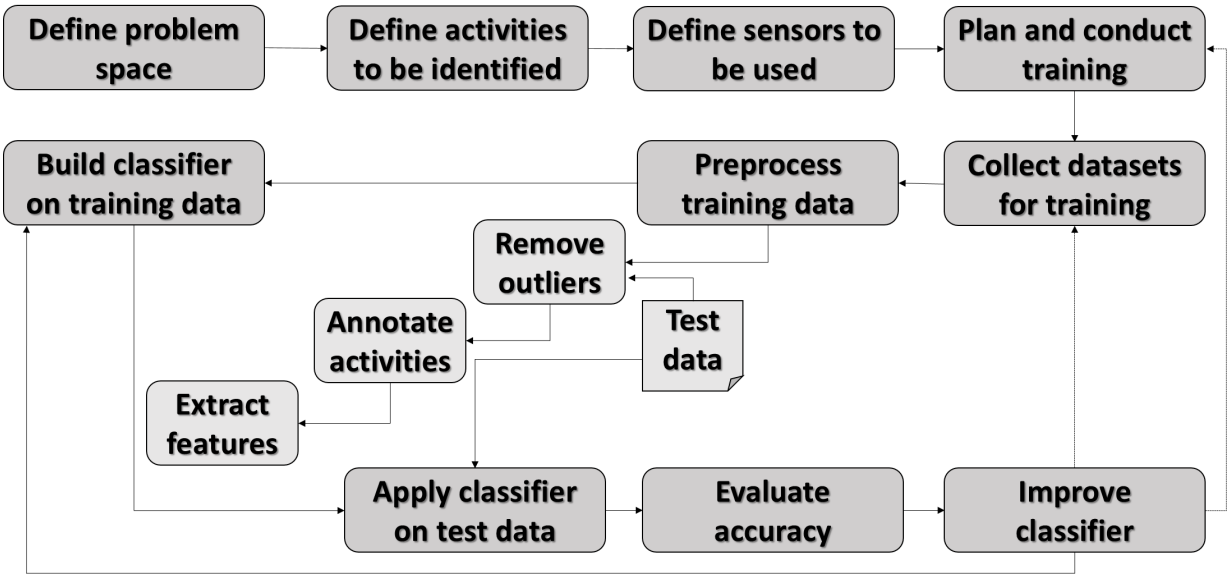

Fig. 3. Generic framework for the analysis of participatory surveillance data.

poses (applies only to training data, since the classifier will ¿predictř the class of the test data) and the extraction of statistical features.

- Build classifier based on training data: amongst the large number of classification algorithm available in the related literature, the optimal one for the particular type of collected data and extracted features should be selected. Each classifier has a set of configuration parameters and a sensitivity analysis of each of them and their influence on the accuracy of the classifier needs to be performed in order to conclude on the most appropriate classifier for the considered experimental settings.
- Apply classifier on test data: having decided on the optimal classifier, it needs to be applied on the collected test data (a posteriori or at runtime depending on the experimental settings). The classifier should be able to identify the class to which each of the test data belongs and decide upon that.
- Evaluate accuracy: the accuracy of the classifier is evaluated against the ground truth, hence the need to properly and accurately annotate both the training and test data.
- Improve classifier: test and training data are of the same type but a lot of irregularities might appear on the test data that might not have been present in the training data. There are many reasons for this, most important of which is the fact that training data users are rarely the same as test data users and thus do not have the same physiological patterns. Therefore, the classifier might not predict the test data as efficiently as expected and further modifications need to be applied on it and accordingly the experiments might need to be performed again.

In what follows we elaborate on the elements of our proposed framework in the context of our participatory surveillance evaluation case-study, i.e. the evacuation exercise.

VI. Evaluation

To validate the feasibility of participatory surveillance and evaluate its efficiency we conducted an evacuation exercise experiment, where the main goal was to establish whether the use of smartphones' sensors as a backup channel for information collection can yield information about users' activities. In this respect, we built a comprehensive dataset for training a statistical classifier and applied this to the test data collected during the experiment to establish possible pattern matches and subsequently analyze the results to augment the design of the assumed participatory surveillance system.

A. Implementation

In order to collect data from the sensors embedded on the smartphones of the users we experimented initially with the Funf open sensing framework [28] running on Android platforms, but we finally opted for a custom built application to avoid the unnecessary complexity in configuring Funf. In particular, Funf allows the developer to set preferences in regards to data collection, such as the types of sensors to be monitored, the duration of the monitoring and recording phase and the interval between two consecutive recording phases. A limitation of this framework is the fact that the actual frequency of data collection, while seemingly subject to a user's preferences, is actually a compromise between the value set by the user and the frequency that Android itself and the corresponding sensors can actually support. Android allows for 4 different rates for data collection from sensors, i.e. *normal, UI, game, fastest*. However, the actual values for these frequencies differ between different sensors, which further complicates matters when it comes to data collected using Funf. For these reasons we built our custom application to collect sensor data, the main distinguishing feature of which

is that it allows explicit setting of data collection frequency per sensor.

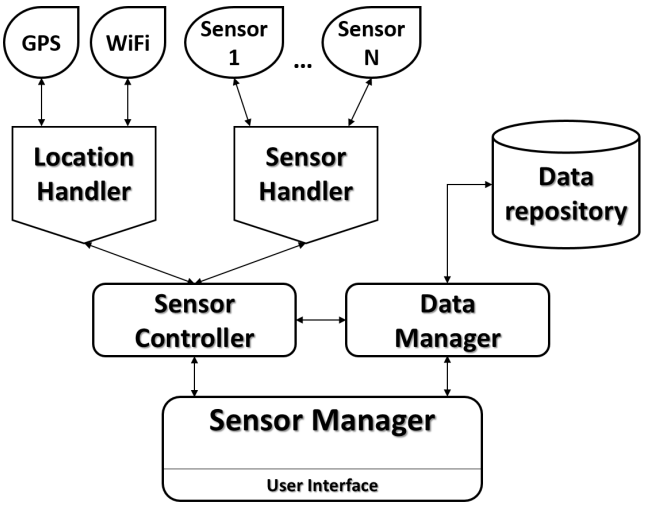

Fig. 4. Architecture of sensor data collection application.

The application we built has a modular and extensible architecture that is shown in Figure 4. The *Sensor Manager* is the main entry point for the app, currently running on Android phones, by means of a dedicated GUI that allows the user to select the sensor the values of which she is interested in recording. Two main elements of the architecture are the *Sensor Controller* and the *Data Manager*. The former interacts with the *Location Handler* to retrieve current location and the *Sensor Handler* to get low level system access to the embedded sensors and retrieve their values, while the latter parses this information into an appropriate data format and stores it to a local *Data Repository*. Moreover, the *Data Manager* supports management of this repository, i.e. search, update, delete data, applying access control policies to avoid unauthorized access to private data and logging every request for data.

B. Experimental setup

The evacuation exercise experiment was conducted in a public building at the JRC premises and it comprised both floors of the building, as well as the parking space, where the actual evacuation meeting points is located. The participatory surveillance exercise was part of a larger experiment that aimed amongst others at evaluating additional techniques such as indoor localization using smartphones and facial recognition through the cameras on the smartphones. For the participatory surveillance tasks, 14 actors were involved: 1 building delegate and 13 regular users carrying their smartphones. The latter had our custom application deployed in order to collect sensor data, e.g. accelerometer, and the users were also instructed at times (via SMS from the control room) to use their smartphones to record video of their surroundings. Users and the building delegate were equipped with different types of smartphones (Samsung Galaxy Nexus, Sony Xperia S, HTC One) to account for the diversity of existing platforms and embedded sensors.

Upon completion of the evacuation exercise we had a total of 13 datasets from a corresponding number of phones that had collected sensor data through our application. Unfortunately, the instability of the application and of the smartphones' platform led to not all phones having recorded data (only 6 out of 13 phones reported worthwhile data). It has to be clarified that the reasons for the erroneous, wrongly timed and limited data collection cannot be pinpointed to a particular event. They can be attributed to users not having activated all services, e.g. location reporting services, networking hindrances due to obstacles or other reasons that collectively prevented timely data reporting, smartphone being overloaded, smartphone battery having been depleted, etc. The major problem that we encountered was the concurrent and synchronized collection of sensor and localization data. This was necessary in order to be able to reason about the sensor data and to pinpoint the location of interesting events. In the future a larger user test base should be considered and an extensive preparation phase prior to the experiments should take place to ensure proper operation. Moreover, test users should not be left to freely interact with the considered applications and services, but instead they should follow a strict script/set of actions in order to ensure that the results we obtain will not be biased by the individual users' attitudes.

The frequency of data collection from the sensors was set to 500Hz (every 2ms), which intuitively is rather high to allow for distinguishing between differences in human activities. Moreover, it reduces the battery level significantly since the device is constantly operational and collecting data. In future experiments we plan to perform a thorough sensitivity analysis of this variable, as well as of other variables in our configuration. The sensors involved in data collection include the triaxial accelerometer, microphone, battery, gravity sensor, gyroscope, light sensor, magnetometer, orientation, WiFi and Bluetooth wireless interfaces and the proximity sensor. While data was collected for all the aforementioned sensors, we focus our analysis here on those collected by accelerometers. The reason behind this decision lies in the fact that in this evacuation exercise, identification of human activities, e.g. walking or running, was the main goal and in this respect the accelerometer has been the sensor most widely used for this purpose in related research works [29].

Moreover, due to the established limitations of the sensing capabilities of modern smartphones and their issues with calibration (see for example [30], [31]), sensors such as the gyroscope, magnetometer and orientation are considered to be of limited accuracy and this placed constraints on their use in our experimental design. An additional reason for not using these sensors was the fact that they are dependent on many external features, e.g. the way the phone is being held, specific rooms/configurations where the experiments took place, so it would have been harder to deduce any useful context from such sensors considering the difficulties in repeating experimental settings and conditions. We therefore opted to

place emphasis on data collection by the triaxial accelerometer present in all current smartphones for the analysis of the data collected in the participatory surveillance project. Evidently, the training dataset can be easily expanded to include data from the other sensors in order to establish whether useful knowledge about the users' status and activities can be extracted from them.

C. Training phase

The focal objective of building a reference training dataset is to establish a solid and wide ground truth to be used for the evaluation of the participatory surveillance experiments, i.e. the identification of distinct human activities. In this respect, we collected training data with a specific focus on data that corresponds to the activities expected to take place during the evacuation exercise; walking, walking up stairs, walking down stairs, standing still. It should be clarified that the more activities we consider, the more accurate the prediction will be since there will be more details and more patterns derived for each of these activities. However, the increase in the number of considered activities brings a corresponding increase in the complexity of the data pre-processing and classification processes. Considering this trade-off we opted for a training dataset containing accelerometer data corresponding to the 4 basic activities mentioned before.

The training dataset contains 5 minutes worth of data collected for each of the aforementioned activities, which were conducted independently and annotated immediately after completion to limit possible annotation errors. Users were asked to retain constant gait and velocity as much as possible. We performed 10 recording sub-sessions, i.e. 40 second sub-sessions, for each activity to eliminate possible user fatigue that would influence data collection. In addition, only 30 seconds were considered out of each sub-session; we trimmed the first and last 5 seconds to omit outliers during initialization and finalization of the activity. Lastly, we used the exact same settings in terms of data collection frequency, smartphone devices and phone placement as in the actual evacuation exercise.

D. Data preprocessing

The collected training data refers to 3 streams of measurements, one for each axis of the accelerometer. They represent continuous discrete samples that need to be pre-processed prior to making any analysis based on them. In accordance with typical activity recognition algorithms, we commence by cleaning the raw data and removing the outliers, followed by the application of a windowing technique to extract groups of data that could potentially expose repetitive activities or tasks with common characteristics, both of which are representative of the majority of human activities. We built a set of custom Java tools that facilitate the automation and efficient preprocessing of the raw data and hence allow to invest more time on the data analysis tasks. All the data are stored as CSV files, the first line of which indicates the type of data.

Removing outliers is a very important task in data preprocessing, since these values could influence the outcome of the analysis process as they are not conforming to the rest of the data and have thus been potentially generated by side activities to the one currently under examination. The most common technique we use to clean the dataset is to remove the influence stemming from the initialization and finalization of activities. We are interested in the execution of the actual activity and therefore we trim the dataset on both ends accordingly. This is similar to the technique used in [7] and we also take into account the risk of significantly reducing the size of the dataset; this is the reason why we collected additional data as previously described. More advanced techniques, e.g. mean, Kalman, particle filters, could also be applied, but since the application of these filters is subject to intense processing requirements, we opted against applying any such technique on the collected accelerometer data.

Subsequently we applied a windowing algorithm to create logical instances, i.e. windows, of the original dataset. The windows are used to reduce the problem space on one hand, but also to assist in grouping similar samples on the other hand. Generally speaking there exist three different windowing techniques, namely sliding, event-based and activity-based windows, applied to raw data for the purpose of recognition of human activities, as discussed in [32]. We used the sliding window technique with overlap 50% over the entire training data population, for a total of over 60,000 samples. Table I illustrates the different values for the window size that we experimented with and the corresponding duration of the window and generated instances in the dataset based on the assumed data collection frequency. We should note that the same techniques for data preprocessing are applied in both the training data as well as the test data during the actual classification process.

TABLE I
WINDOW SIZE, DURATION AND SIZE OF INSTANCES IN DATASETS

Window size	Duration	Instances in dataset
64	1.28s	1875
128	2.56s	937
256	5.12s	468
512	10.24s	232

The sampling frequency of the original training dataset is too high for individual samples to exhibit any interesting properties, especially in regards to human activities that normally have a duration lasting at least a few seconds. To alleviate this concern, we extract features over the different instances (windows) of the data. The notion of features refers to statistical properties of the windows of data and provides some qualitative or quantitative information on them. We examined time domain and frequency domain features in the related literature and we selected a total of 105 features to compute for all the data contained in each of the windows in the dataset. Out of these 105 features, 75 were in the time domain and 30 in the frequency domain. The latter necessitated

that we first apply a Fast Fourier Transform (FFT) over the windows' data and then calculate the corresponding features. Table II summarizes the selected features that were applied on the accelerometer values for each of the three axes, the magnitude of acceleration and the tilt on all three axes.

TABLE II
TIME AND FREQUENCY DOMAIN FEATURES FOR CLASSIFICATION

Time domain	Frequency domain
Mean	Mean
Median	Median
Minimum	Minimum
Maximum	Maximum
Harmonic mean	Spectral energy
Geometric mean	Entropy
Pearson's correlation	Pearson's correlation
Variance	
Standard deviation	
Root mean square	
Covariance	

Cyclic patterns of human activities can be easier observed using frequency domain features and this is the main reason for their significance. Indicatively, Figure 5 depicts the mean FFT for the 3 axes of acceleration; cyclic, identifiable patterns can be clearly seen in the activities involving stairs as expected.

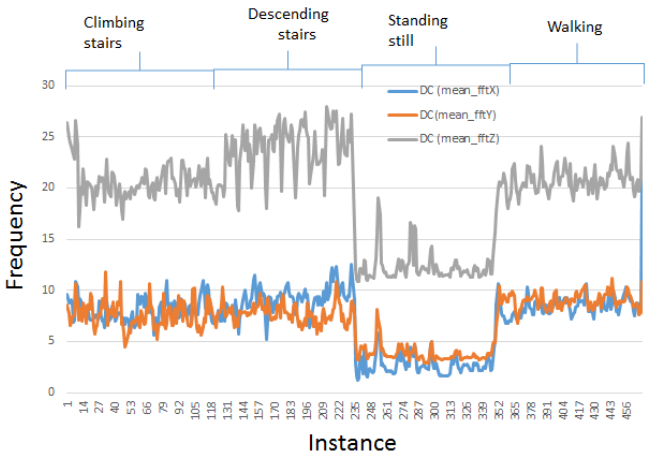

Fig. 5. Mean FFT of acceleration in the 3 axes and the activities that it corresponds to.

E. Data classification

Having completed the aforementioned procedures and in accordance to our generic framework, the next step in the process involves training the classifier in order for it to be able to correctly analyze the collected data, but also to be able to respond to prospective queries on how new data should be classified, i.e. annotated, in line with previous data. We used the Weka toolkit [33] to experiment with a variety of machine learning algorithms, namely KStar, C4.5 decision tree, Bayes (Bayes Network with K2 search algorithm), Support Vector Machines (SVM), Sequential Minimal Optimization (SMO), k-NN, MultiClass (meta-classifier) and MultiLayer Perceptron Neural Network(MLP).

We indicatively present in the following results on the training of the classifiers based on training data coming from a Sony Ericsson phone, where the data collection frequency was set to be 500, the window size 256 with 50% overlap, the user conducted all 4 activities and she did so holding the phone in her hand at the waist level. In order to test the accuracy of the various classifiers, we used Weka and performed a comparative analysis of the aforementioned 8 algorithms using cross-validation with 10 folds and 10 iterations using all the training data. The annotated test data were used for the evaluation and that is where the corresponding results refer to. To gain a level of statistical confidence in the results obtained by Weka, we applied the well-known statistical hypothesis test Students' T-Test, with a requested confidence of 0.05 (indicates statistical difference threshold when performing pairwise comparison between schemes), and managed to acquire measurements for a variety of aspects regarding accuracy and the performance of the classifiers accordingly. Results regarding these classifiers are presented in Table III and discussed in the following.

All classifiers exhibited very high accuracy rates, reaching up to 99.5%. It becomes therefore evident that the use of just the accelerometer in detecting and distinguishing between human activities yields very promising results. Additionally, it exposes the privacy risks involved in participatory sensing, since it illustrates the level of knowledge that can be gained with the use of a non-intrusive sensor on board a smartphone.

F. Analysis of results

Table III clearly highlights the optimal performance of the SVM classifier over all others, with the C4.5 one a close second. SVM has an accuracy level of 99.5%, whereas C4.5 98.8% and the worst performing classifiers are the Bayes Network and KStar with accuracies of 85% and 86.6%, respectively. It is interesting to examine this observation in light of the mean absolute and root mean squared error metrics. A simplistic analysis would expect the Bayes Network and KStar classifiers to have the largest degrees of errors, due to their low accuracy. However, while Bayes Networks and KStar do not indeed fare well, it is actually the SMO and MLP classifiers that exhibit the largest mean errors and root mean squared errors despite high accuracy levels of 97.01% and 97% respectively. The reason for this is the fact that while the latter classifiers managed to correctly classify a larger number of instances, the misclassifications were so big that they succeeded in significantly increasing the mean errors. This observation is consistent with our belief that evaluation of a machine learning classifier is not a simple process of computing a couple of metrics, but rather an extensive procedure where a series of quantitative metrics should be taken into account and considered in parallel, while additionally one should not neglect qualitative analysis, as discussed later.

Two very important metrics in assessing classifiers are the precision and the recall (borrowed from the field of Information Retrieval - IR). Precision measures successful assignments to a class over all assignments to that class (including incorrect ones) and in this respect it refers to the

TABLE III
TIME AND FREQUENCY DOMAIN FEATURES FOR CLASSIFICATION

Classifier	C4.5	Bayes	KStar	SVM	SMO	kNN	MLP	MultiClass
Evaluation metric								
Accuracy	98.8054	85.0457	86.5528	99.4976	97.1061	91.2616	97.0416	96.8560
Incorrect classifications	1.1946%	14.9543%	13.4472%	0.5024%	2.8939%	8.7384%	2.9584%	3.1440%
Mean absolute error	0.0060	0.0748	0.0659	0.0025	0.2529	0.0673	0.3044	0.0208
Root mean squared error	0.0638	0.2642	0.2472	0.0345	0.3159	0.1814	0.3523	0.1164
IR Precision	1.000	0.7704	0.7720	0.9976	0.9516	0.8294	0.9574	0.9842
IR Recall	0.9850	0.7762	0.7814	0.9974	0.9799	0.8767	0.9698	0.9497
F-measure	0.9923	0.7715	0.7745	0.9975	0.9652	0.8501	0.9625	0.9659
Area under ROC	0.9925	0.9461	0.9585	0.9983	0.9879	0.9741	0.9999	0.9942
KB mean information	1.9712	1.6417	1.6807	1.9879	0.9803	1.7244	0.6418	1.9052
Kappa statistic	0.9841	0.8006	0.8207	0.9933	0.9614	0.8835	0.9606	0.9581
Elapsed time training	0.0902s	0.1299s	0.0004s	1.1515s	0.1447s	0.0006s	176.3556s	0.4553s
Elapsed time testing	0.0069s	0.0244s	10.3873s	0.1052s	0.0079s	0.1009s	0.1494s	0.0578s

fraction of classified instances that are relevant, i.e. correct. Conversely, recall refers to the fraction of relevant instances that have been classified. Therefore, high recall values indicate that the classifier was successful in classifying correctly most of the instances, whereas high precision means that the classifier performed more correct classifications than incorrect ones. In our experiments, it is the C4.5 classifier that has the best precision with a value of 1.0, followed by the SVM with a value of 0.99, while the Bayes Network and KStar classifiers have the lowest precision with value of 0.77 for both of them. In general, precision follows the same pattern as accuracy, which was to be expected since these two metrics are conceptually close. Similar results (top two algorithms being SVM and C4.5 and lower two Bayes Network and KStar) can be seen for the recall metric, albeit with more distinguish values. Interestingly enough, while the MultiClass classifier had the third best precision at 0.98, it is the SMO classifier that has the third best recall at 0.98. This indicates a different performance between these two, where MultiClass is better at classifying more instances correctly than incorrectly and SMO outperforms MultiClass in performing more correct classifications. Nevertheless, SMO had a much higher root mean squared error than MultiClass, so in general one can expect better performance of the latter.

Another very important metric is the F-measure that combines both precision and recall (also known as F1 score or F1 measure since precision and recall are evenly weighted). It is actually the harmonic mean of precision and recall and is used to express the accuracy of the classification process and is widely considered to be more useful than the percent of correct classifications as expressed by the accuracy metric. According to the F-measure, the SVM and C4.5 perform the best, while the Bayes Network and KStar the worst. Moreover, the area under ROC (receiver operating characteristic) metric has also proven to be extremely useful in evaluating classification algorithms [34], although its value has been recently heavily criticized and thus undermined, e.g. in [35]. The area under the ROC curve is equal to 1 for a perfect classification and drops as classification quality drops. In our experiments MLP performs the best in terms of the area under ROC with a value very close to 1, followed by SVM and MultiClass, while Bayes Networks and KStar perform the worst. It has to be clarified nonetheless that even the worst performing classifier, i.e. Bayes Network, has a value equal to 0.95 that broadly speaking is very good for classification purposes.

Other metrics that we considered in order to compare the performance of the considered classifiers include the Kappa statistic and the KB mean information. Kappa statistic is used to indicate the agreement of prediction compared to the ground truth and is important since it is a probabilistic value that takes into account not only the comparison to the ground truth, but also the probability that a correct assignment to a class was by chance. As before the SVM and C4.5 classifiers were the ones with higher Kappa statistic value at 0.99 and 0.98 respectively (the higher the value, the better matching the agreement), while KStar and Bayes Network had the lowest values, 0.82 and 0.8 respectively. Kononenko and Bratko [36] introduced an information-based evaluation criterion for each classifier's performance, which excludes prior class probabilities and thus assesses better the performance of the classifier under uncertain conditions. Once again SVM and C4.5 had the highest performance in regards to this metric, but surprisingly MLP did not perform well enough. In our view, the reason is based on the construction of the MLP construction network that inherently requires knowledge of prior class probabilities (back-propagation error correction is at its core), so when these probabilities are excluded its performance is bound to be reduced.

The last metric that we considered was the aspect of time. In particular, we examined the time required to train the classifiers and the time required for them to perform classifications over the test data. Since the classifiers were trained and tested on the same sets of data the values obtained for these metrics are directly comparable. MLP, which is the most complex of the considered classifiers, requires the most time for training, averaging 176.35 seconds. This is however not reflected in the testing phase that only takes 0.15 seconds. The fastest classifier to train is KStar followed closely by k-NN, whereas apart from the MLP classifier, the SVM one at 1.15 seconds is also relatively slow to train. Conversely, when it comes to testing times SVM is quite fast at 0.1 seconds but the fastest ones are C4.5 and SMO. The slowest classifier for

testing was KStar with the remaining algorithms exhibiting small variance in their values.

```
Scheme:weka.classifiers.functions.LibSVM -S 0 -K 0 -D 3 -G 0.0 -R 0.0
Relation:     training_data_A_H_XX_50_SE_features_256 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1
Instances:    468
Attributes:   107
[list of attributes omitted]
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 0.62 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         467               99.7863 %
Incorrectly Classified Instances         1                0.2137 %
Kappa statistic                          0.9972
Mean absolute error                      0.0011
Root mean squared error                  0.0327
Relative absolute error                  0.2849 %
Root relative squared error              7.5481 %
Total Number of Instances              468

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.003      0.992     1         0.996      0.999    climbing_stairs
                 0.991     0          1         0.991     0.996      0.996    descending_stairs
                 1         0          1         1         1          1        standing_still
                 1         0          1         1         1          1        walking
Weighted Avg.    0.998     0.001      0.998     0.998     0.998      0.999

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 117   0   0   0 |   a = climbing_stairs
   1 116   0   0 |   b = descending_stairs
   0   0 117   0 |   c = standing_still
   0   0   0 117 |   d = walking
```

Fig. 6. Confusion matrix for the SVM classifier.

```
J48 unpruned tree
------------------

instance <= 234
|   instance <= 117: climbing_stairs (117.0)
|   instance > 117: descending_stairs (117.0)
instance > 234
|   var_accZ <= 0.9846: standing_still (117.0)
|   var_accZ > 0.9846: walking (117.0)

Number of Leaves  :     4

Size of the tree :      7


Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         464               99.1453 %
Incorrectly Classified Instances         4                0.8547 %
Kappa statistic                          0.9886
Mean absolute error                      0.0043
Root mean squared error                  0.0654
Relative absolute error                  1.1395 %
Root relative squared error             15.0962 %
Total Number of Instances              468

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.991     0          1         0.991     0.996      0.996    climbing_stairs
                 0.991     0.003      0.991     0.991     0.991      0.994    descending_stairs
                 0.991     0.006      0.983     0.991     0.987      0.993    standing_still
                 0.991     0.003      0.991     0.991     0.991      0.994    walking
Weighted Avg.    0.991     0.003      0.991     0.991     0.991      0.994

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 116   1   0   0 |   a = climbing_stairs
   0 116   1   0 |   b = descending_stairs
   0   0 116   1 |   c = standing_still
   0   0   1 116 |   d = walking
```

Fig. 7. Confusion matrix for the C4.5 classifier.

It is clear that there are tradeoffs to be considered when choosing the optimal classifier for participatory surveillance needs such as identifying human activities. We need to consider accuracy, precision, recall, as well as the overhead in terms of time for each of the classifiers since they will need to be considered in real time operation. The training phase will only occur once, so long timespans for this phase can be sidestepped, but long times for testing can be used to exclude certain classifiers from our candidates' list. Evidently, quantitative results as those previously presented are important, since they provide a thorough evaluation of the performance of the different classifiers in regards to a variety of aspects. It is however equally important to be able to qualitatively analyze the classification process and in particular to be able to analyze why classification errors occur. The best way to do this is by checking the confusion matrix (also known as contingency table) of the classification process that represent the classification results versus the ground truth. Indicatively, Figure 6 shows the confusion matrix for the SVM classifier and Figure 7 for the C4.5 one. Confusion matrices are important because they allow us to diagnose which classes were confused to each other and therefore be able to draw conclusions as to why this occurred in the first place.

Based on the aforementioned extensive analysis and evaluation of the considered classifiers we came to the conclusion that the most suitable ones for participatory surveillance needs include the SVM and C4.5 one. They exhibited the optimal balance between performance, accuracy and quality of classification.

VII. Conclusion

In this paper we presented our work on designing and developing a solution for participatory surveillance. We aim at involving end users in the tasks related to security and surveillance and thus on one hand assist and promote the overall perceived level of safety, while on the other hand promoting users' sense of contribution and participation in the society and hence their awareness. By utilizing the numerous sensors on smartphones that are nowadays ubiquitous we postulate that significant information regarding critical, security-related events can be inferred. As a proof of concept, we built a system to collect such data from users in the context of an emergency evacuation exercise and we presented here relevant results on the use of this data. By using just one sensor, namely the accelerometer, very high levels of accuracy in predicting users' activities were reached. In our view, this validates the great potential that exists in the field of participatory surveillance, in particular for the management of emergency/crisis events. Even with quite a few limitations that we encountered in our study, e.g. sensors accuracy or user participation, and with the limited amount of collected data, intelligence on the different activities of users was deducible.

Based on the collected results and our analysis, we are confident that with the integration of additional sensors, as well as with the collection of a far more detailed reference dataset, we would definitely be able to discern between distinct human activities at a much greater level of detail and with quantifiable assessment metrics. These aspects are among the ones we plan

to investigate further in the future. To test the feasibility of our machine learning approach on participatory surveillance data we did not check against all possible testing conditions; this extensive sensitivity analysis is nonetheless the focus of our ongoing work. We are also planning work on examining the potential benefits that might arise from exploiting additional sensors such as the magnetometer, gyroscope, etc.

The usefulness of participatory surveillance is extremely high if one considers the fact that such a framework could for example allow groups of rescuers to gain access to information about the current and ongoing status and activities of people inside a building, e.g. static user for a long time or user suffering a physical shock. The analysis of the results and the possibility of detecting with high accuracy the class of previously unclassified data has highlighted the great potential of participatory surveillance systems. However, it has also exposed the great privacy risks regarding users sharing data from their smartphones from such systems. We believe that the use of additional sensors and the information fusion emerging from the use of multiple sensors will exacerbate these privacy risks and allow for more accurate detection of the users' activities, as well as the context of her surroundings. Typical examples of such risks reported in the literature include the possibility to infer the PIN of users on smartphones or the password that they type using accelerometers and gyroscopes [37]. We therefore also plan to examine the risks involved from the potential sharing of data from a variety of sensors and not only the accelerometer.

REFERENCES

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *In: Workshop on World-Sensor-Web (WSWd'06): Mobile Device Centric Sensor Networks and Applications*, 2006, pp. 117–134.

[2] D. Estrin, "Participatory sensing: applications and architecture [internet predictions]," *Internet Computing, IEEE*, vol. 14, no. 1, pp. 12–42, 2010. doi: 10.1109/MIC.2010.12

[3] K. Shilton, "Participatory sensing: Building empowering surveillance," *Surveillance & Society*, vol. 8, no. 2, pp. 131–150, 2010.

[4] Z. Dong, B. Lu, L. He, P. Cheng, Y. Gu, and L. Fang, "Exploring smartphone-based participatory computing to improve pervasive surveillance," in *11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '13. ACM, 2013. doi: 10.1145/2517351.2517388. ISBN 978-1-4503-2027-6 pp. 69:1–69:2. [Online]. Available: http://doi.acm.org/10.1145/2517351.2517388

[5] F. Coudert, M. Gemo, L. Beslay, and F. Andritsos, "Pervasive monitoring: Appreciating citizen's surveillance as digital evidence in legal proceedings," in *Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th Intl Conference on*, 2011. doi: 10.1049/ic.2011.0130 pp. 1–6.

[6] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 140–150, 2010. doi: 10.1109/MCOM.2010.5560598

[7] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, ser. LNCS, A. Ferscha and F. Mattern, Eds. Springer, 2004, vol. 3001, pp. 1–17. ISBN 978-3-540-21835-7. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-24646-6_1

[8] T. Huynh and B. Schiele, "Analyzing features for activity recognition," in *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-aware Services: Usages and Technologies*, ser. sOc-EUSAI '05. New York, NY, USA: ACM, 2005. doi: 10.1145/1107548.1107591. ISBN 1-59593-304-2 pp. 159–163. [Online]. Available: http://doi.acm.org/10.1145/1107548.1107591

[9] T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part II: Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, ser. IWANN '09. Berlin, Heidelberg: Springer-Verlag, 2009. doi: 10.1007/978-3-642-02481-8_120. ISBN 978-3-642-02480-1 pp. 796–799. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02481-8_120

[10] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Medusa: A programming framework for crowd-sensing applications," in *10th Intl Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12. ACM, 2012. doi: 10.1145/2307636.2307668. ISBN 978-1-4503-1301-8 pp. 337–350. [Online]. Available: http://doi.acm.org/10.1145/2307636.2307668

[11] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "Peir, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1555816.1555823. ISBN 978-1-60558-566-6 pp. 55–68. [Online]. Available: http://doi.acm.org/10.1145/1555816.1555823

[12] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "Greengps: A participatory sensing fuel-efficient maps application," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1814433.1814450. ISBN 978-1-60558-985-5 pp. 151–164. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814450

[13] M. Wisniewski, G. Demartini, A. Malatras, and P. Cudré-Mauroux, "Noizcrowd: A crowd-based data gathering and management system for noise level data," in *Mobile Web Information Systems*, ser. LNCS, F. Daniel, G. Papadopoulos, and P. Thiran, Eds. Springer, 2013, vol. 8093, pp. 172–186. ISBN 978-3-642-40275-3

[14] K. Shilton, "Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection," *Comm. of the ACM*, vol. 52, no. 11, pp. 48–53, Nov. 2009. doi: 10.1145/1592761.1592778. [Online]. Available: http://doi.acm.org/10.1145/1592761.1592778

[15] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *Communications Magazine, IEEE*, vol. 49, no. 11, pp. 32–39, 2011. doi: 10.1109/MCOM.2011.6069707

[16] H. Keval and M. A. Sasse, "To catch a thief – you need at least 8 frames per second: The impact of frame rates on user performance in a cctv detection task," in *Proceedings of the 16th ACM International Conference on Multimedia*, ser. MM '08. New York, NY, USA: ACM, 2008. doi: 10.1145/1459359.1459527. ISBN 978-1-60558-303-7 pp. 941–944. [Online]. Available: http://doi.acm.org/10.1145/1459359.1459527

[17] A. Ito, A. Aiba, A. Ito, and S. Makino, "Detection of abnormal sound using multi-stage gmm for surveillance microphone," in *Information Assurance and Security, 2009. IAS '09. Fifth International Conference on*, vol. 1, Aug 2009. doi: 10.1109/IAS.2009.160 pp. 733–736.

[18] J. A. Hanson, K. L. McLaughlin, and T. J. Sereno, "A flexible data fusion architecture for persistent surveillance using ultra-low-power wireless sensor networks," pp. 80 470M–80 470M–12, 2011. [Online]. Available: http://dx.doi.org/10.1117/12.883280

[19] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-efficient surveillance system using wireless sensor networks," in *Proceedings of the 2Nd International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '04. New York, NY, USA: ACM, 2004. doi: 10.1145/990064.990096. ISBN 1-58113-793-1 pp. 270–283. [Online]. Available: http://doi.acm.org/10.1145/990064.990096

[20] T. Monahan and J. T. Mokos, "Crowdsourcing urban surveillance: The development of homeland security markets for environmental sensor networks," *Geoforum*, vol. 49, no. 0, pp. 279 – 288, 2013. doi: http://dx.doi.org/10.1016/j.geoforum.2013.02.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0016718513000341

[21] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, ser. Ubicomp '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1864349.1864355. ISBN 978-1-60558-843-8 pp. 33–36. [Online]. Available: http://doi.acm.org/10.1145/1864349.1864355

[22] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, P. Floréen, A. Krüger, and M. Spasojevic, Eds. Springer Berlin Heidelberg, 2010, vol. 6030, pp. 138–155. ISBN 978-3-642-12653-6. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-12654-3_9

[23] I. Martí, L. Rodríguez, M. Benedito, S. Trilles, A. Beltrán, L. Díaz, and J. Huerta, "Mobile application for noise pollution monitoring through gamification techniques," in *Entertainment Computing - ICEC 2012*, ser. Lecture Notes in Computer Science, M. Herrlich, R. Malaka, and M. Masuch, Eds. Springer Berlin Heidelberg, 2012, vol. 7522, pp. 562–571. ISBN 978-3-642-33541-9. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33542-6_74

[24] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, Nov. 2011. doi: 10.1016/j.jss.2011.06.073. [Online]. Available: http://dx.doi.org/10.1016/j.jss.2011.06.073

[25] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Computer Communications*, vol. 33, no. 11, pp. 1266 – 1280, 2010. doi: http://dx.doi.org/10.1016/j.comcom.2009.08.012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366409002448

[26] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 4, pp. 907–920, April 2014. doi: 10.1109/TMC.2013.28

[27] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN 0387310738

[28] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland, "Social fmri: Investigating and shaping social mechanisms in the real world," *Pervasive and Mobile Computing*, vol. 7, no. 6, pp. 643 – 659, 2011. doi: http://dx.doi.org/10.1016/j.pmcj.2011.09.004 The Ninth Annual {IEEE} International Conference on Pervasive Computing and Communications (PerCom 2011). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119211001246

[29] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 33:1–33:33, Jan. 2014. doi: 10.1145/2499621. [Online]. Available: http://doi.acm.org/10.1145/2499621

[30] C. Barthold, K. Subbu, and R. Dantu, "Evaluation of gyroscope-embedded mobile phones," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE Intl Conference on*, Oct 2011. doi: 10.1109/ICSMC.2011.6083905. ISSN 1062-922X pp. 1632–1638.

[31] Z. Wu, Y. Wu, X. Hu, and M. Wu, "Calibration of three-axis magnetometer using stretching particle swarm optimization algorithm," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 2, pp. 281–292, Feb 2013. doi: 10.1109/TIM.2012.2214951

[32] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensorsÛa review of classification techniques," *Physiological Measurement*, vol. 30, no. 4, p. R1, 2009. [Online]. Available: http://stacks.iop.org/0967-3334/30/i=4/a=R01

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009. doi: 10.1145/1656274.1656278. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656278

[34] K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning," in *Proceedings of the Sixth International Workshop on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1989. ISBN 1-55860-036-1 pp. 160–163. [Online]. Available: http://dl.acm.org/citation.cfm?id=102118.102172

[35] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the roc curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, Oct. 2009. doi: 10.1007/s10994-009-5119-5. [Online]. Available: http://dx.doi.org/10.1007/s10994-009-5119-5

[36] I. Kononenko and I. Bratko, "Information-based evaluation criterion for classifier's performance," *Machine Learning*, vol. 6, no. 1, pp. 67–80, Jan. 1991. doi: 10.1023/A:1022642017308. [Online]. Available: http://dx.doi.org/10.1023/A:1022642017308

[37] Z. Xu, K. Bai, and S. Zhu, "Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors," in *5th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WISEC '12. ACM, 2012. doi: 10.1145/2185448.2185465. ISBN 978-1-4503-1265-3 pp. 113–124. [Online]. Available: http://doi.acm.org/10.1145/2185448.2185465