Activity Recognition from Mobile Phone Data: State of the Art, Prospects and Open Problems

DRISS CHOUJAA
and
NARANKER DULAY
Imperial College London

Activity recognition aims to automatically determine what people do from a series of observations. With several billion subscribers and multiple sensors, the mobile phone is an obvious opportunity for activity recognition. We give an overview of mobile-phone-based activity recognition along two directions – recognised activities and inference techniques. We show that existing systems can be classified in two categories depending on whether they are location- or motion-driven. A new generation of smartphones is inspiring many research efforts. However, the current framework for mobile-phone-based activity recognition has reached certain limits. We anticipate changes in the field and identify obstacles to further development.

Categories and Subject Descriptors: A.1 [**Introductory and survey**]: ; C.5.3 [**Microcomputers**]: Portable devices (e.g., laptops, personal digital assistants); I.2.6 [**Learning**]:

General Terms: Algorithms, Design, Experimentation, Human Factors

Additional Key Words and Phrases: Activity recognition, Mobile phone, Cellphone

1. INTRODUCTION

The layman usually has a limited view of what can be achieved through the analysis of mobile phone data. Mobile phone users often express certain privacy concerns over service providers being able to inspect their phone calls or text messages without their consent. Mobile phone data analysis can actually do much more. With 4.1 billion subscribers throughout the world [Tryhorn 2009], the mobile phone is creating new opportunities to collect, understand and utilise information about human behaviour. Through their many sensors, today's mobile phones can unobtrusively record not only communication but also location and co-presence of individuals throughout the day. By applying machine learning techniques to this data, systems can determine what the user is doing at any point in time with reasonable accuracy. Travel, for example, can be recognised from fluctuations in cellular signals [Anderson and Muller 2006a] and time at work can be estimated by detecting colleagues' Bluetooth-enabled mobile phones or desktop computers [Eagle and Pentland 2006]. Furthermore, the networked nature and computing capabilities of mobile phones make them much more than just logging devices and opens the way to new applications. In particular, the analysis of human behaviour from mobile phone data is

This research was undertaken as part of the Cityware project (www.cityware.org.uk) and supported by UK EPSRC research grant EP/C547705/1.
Authors' addresses: D. Choujaa, Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom; email: dfc05@doc.ic.ac.uk; N. Dulay, Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom; email: nd@doc.ic.ac.uk.

expected to impact three areas profoundly:

—**Marketing.** Mobile phones provide personal, anywhere and anytime access to the consumer. Text-message marketing is already being used to send people text coupons when they are shopping in the vicinity of certain businesses [Cohen 2009]. Companies such as McDonald's, Burger King, Procter & Gamble, General Motors and CBS have also been involved in opt-in text message campaigns for several years [The Associated Press 2006]. By automatically identifying users' weekly schedule, lifestyle, journeys and acquaintances, mobile phone data analysis allows fine-grained customer-profiling and targeted advertisement.

—**Security.** Defense and homeland security departments have been among the first to explore phone and location data [Shachtman 2003]. The ability to determine the behaviour of mobile phone users continuously indeed offers new possibilities for investigators. Cases of successful use of mobile phones to track suspects, victims and witnesses in police investigations already abound around the world [Barnard 2009]. Mobile phone data analysis could provide law enforcement agencies with continuous access to the activities and co-presence of suspicious individuals.

—**Social networking.** Mobile social networking services have expanded in recent years. For example, the micro-blogging service Twitter [Twitter, Inc. 2006], created in 2006, enables its users to send and receive short updates about their activities via a website, text messages or external applications. Twitter's website passed 50 million unique worldwide visitors in July 2009 [Rao 2009]. Other major Internet companies such as Facebook [Facebook, Inc. 2009] and Google [Google, Inc. 2009a] are entering the market of mobile social networking. For example, Google Latitude [Google, Inc. 2009b] allows its users to share their location with others, as estimated by their mobile phone. In the future, diaries and blogs could be generated automatically by recognising a mobile phone user's activities throughout the day.

Research in automated human behaviour analysis has also been driven by other applications which will have a significant social and economic impact:

—**Logistics.** Production can be increased by identifying a worker's actions and delivering just-in-time instructions about what has to be done next [Antifakos et al. 2002]. Boeing has pioneered activity-aware assembly [Barfield and Caudell 2001] while wearable activity recognition systems have been tested on Skoda's car manufacturing process [Stiefmeier et al. 2008]. In office environments, meetings and conferences can be detected with high accuracy from audio and video [Oliver and Horvitz 2004]. Mobile phones can then be turned to silent mode in a meeting or set to vibrate for a longer period of time when the carrier is walking outdoors, to ensure that call notifications are not missed between strides [Anderson and Muller 2006b]. Recognising workers' activities provides a clear picture of time utilisation and can help make working environments more efficient [Osmani et al. 2008].

—**Healthcare.** Recognising simple activities such as *walking*, *running* and staying *stationary* can already help calculate a mobile phone user's energy expenditure through daily step counts [Sohn et al. 2006]. By also considering the user's location and encounters, one can easily imagine estimating exposure to pollutants, noise and infections in the near future. Automatically-generated behaviour logs can serve as memory aids for aging populations [Vemuri and Bender 2004] and help doctors monitor their patients continuously. Mobile activity-aware devices can also assist patients suffering from mild cognitive impairment in public transportation [Liao et al. 2007] and help patients suffering from diabetes to maintain stable blood glucose levels [Preuveneers and Berbers 2008].

—**Policy and planning.** Since the 1970s, policy makers, urban planners and economic analysts have shown a growing interest in human behaviour data. In 2003, after years of study, the U.S. Bureau of Labor Statistics launched the first annual American Time Use Survey (ATUS) [U.S. Bureau of Labor Statistics 2009]. Each year, randomly selected individuals are asked to fill out a one-day diary with his or her activities in 15-minute intervals. Whilst surveys of this kind provide useful information, they place a heavy burden on respondents and are expensive to execute. This naturally limits their scale, duration and frequency. For cost reasons, the size of the surveyed population was reduced from 40,500 individuals in 2003 to approximately 26,000 in the following years [Kimmel 2008]. Mobile phones allow the collection of behaviour data on samples larger than those of surveys by several orders of magnitude. Importantly, the automatic collection of mobile phone data circumvents human reports, giving direct and real-time access to people's activities.

—**Energy saving.** There is growing awareness of environmental issues throughout the world. Predicting when a house's inhabitants will return from work or school helps determine an optimal heating strategy [Gupta et al. 2009]. The energy footprints of home and office buildings can also be reduced by automatically switching on electronic devices when they are likely to be used and switching them off otherwise.

—**Entertainment.** Activity-aware music players aim to play music that fits the user's current level of activity [Dornbush et al. 2007], [Park et al. 2006], [Elliott and Tomlinson 2006]. Industry leaders such as Nike and Apple have integrated awareness of human activities into portable music players by equipping running shoes with sensors and wireless interfaces [Apple, Inc. 2006]. Since recent mobile phones are also portable music player, mobile-phone-based activity recognition could be utilised to select songs based on their tempo [Györbíró et al. 2009]. Mobile games are also beginning to sense players' locations and actions [Hewlett-Packard Development Company 2008], [Groundspeak 2000].

All the applications mentioned above assume that a system is capable of understanding automatically what a user is doing. Implementing this capability is the objective of activity recognition[1]. In order to fulfill its objective, a mobile-phone-based activity recognition system records radio signals and other inputs from mobile phone sensors, processes this data to extract useful information and provides the result to a machine learning classifier for training. Usually, training data also includes the user's true activity when data was collected, as provided by the user

[1]other names plan/goal recognition

himself. The trained classifier can then be utilised to determine the user's activity in new situations from previously-unseen sensor data.

The mobile phone is an obvious opportunity for activity recognition. People carry their mobile phone with them for communication purposes. Therefore, the many sensors embedded in this device can capture data about its user's activities unobtrusively throughout the day. However, mobile-phone-based activity recognition also faces specific challenges which impede its large-scale deployment. The objective of this survey is twofold. First, it provides an entry point for those who may be interested in learning about activity recognition from mobile phone data. The increasing programmability of mobile phones indeed makes these platforms accessible to a larger audience than ever before. We give an overview of the task along two directions – recognised activities and inference techniques. We also show that existing systems can be classified in two main categories depending on whether they are location- or motion-driven. A new generation of smartphones – popular through the commercial success of the Apple iPhone [Apple, Inc. 2009] – offers advanced sensing, processing and connectivity capabilities which are inspiring many research efforts in both industry and academia. However, the current framework for mobile-phone-based activity recognition seems to have reached certain limits. Recent systems achieve relatively high accuracy but focus on a very small set of human activities. We therefore believe that it is particularly timely to identify challenges in this domain and consider the next directions of research. In order to serve this second objective, we anticipate forthcoming changes in the field and determine obstacles to further development.

We present in this paper the first literature survey on activity recognition from mobile phone data. The next section provides an overview of the task along two directions – inferred activities and inference techniques. In Section 3, we review existing mobile-phone-based activity recognition systems and show that they can be categorised in either of two categories, depending on whether they are location- or motion-driven. We then examine prospects in mobile-phone based activity recognition by anticipating technological evolutions of the mobile phone platform. Lastly, we identify in Section 5 seven research challenges and open problems in the task.

2. OVERVIEW OF THE TASK

In general terms, activity recognition consists in associating sensor readings and other inputs to a label taken from a set of distinguishable activities. The task therefore involves (i) determining a set of activity labels and (ii) assigning sensor readings and other inputs to the appropriate activity labels.

2.1 Activity Labels

Activity labels are the output of an activity recognition system. Defining a set of activity labels therefore sets a target which directs the entire design of a system from sensing to modelling. In this section, we review activity labels assigned by current activity recognition systems and classify them into two categories.

2.1.1 *Defining Activity Labels.* The Merriam-Webster dictionary [Merriam-Webster, Inc. 2009] defines an activity as (i) *the quality or state of being active* or (ii) *a vigorous or energetic action.* Activity thus encompasses both state and energy. In mobile and ubiquitous computing, authors generally avoid defining activities out of context. Indeed, most often, the set of activity labels considered is constrained by the sensors at hand and depends on the final application. A fitness monitoring system will define *stationary*, *walking* and *running* as activities whereas an automatic diary application will use just one *travelling* activity and divide the *stationary* state into several depending on the location of the user, for instance, *staying at home* and *working in the office*. For certain applications, sets of activities can be borrowed from other disciplines. For instance, Activities of Daily Living (ADLs) [Katz et al. 1963], [Katz 1983] have been shown to be relevant in many healthcare situations and several activity recognition systems have attempted to address them specifically [Tapia et al. 2004], [Park and Kautz 2008], [Philipose et al. 2004]. Examples of ADLs are *eating*, *dressing*, *getting into or out of a bed or chair*, *taking a bath or shower* and *using the toilet*. Historically, mobile-phone-based systems have focused on a smaller set of activities because of the limited sensing capabilities of early mobile phones. In many cases, human activity is approximated with two proxies, namely *location* and *motion*, which correspond to two orthogonal types of activity recognition.

2.1.2 *Locational versus Motional activity recognition.* The best way of introducing *locational activity recognition* may be to compare it to the better known and closely related *localisation* task [Hightower and Borriello 2001]. In the latter, the objective is to estimate the position of a device anywhere as accurately as possible using techniques such as multilateration or assisted GPS. In contrast, locational activity recognition is interested in the meaning of the user's location rather than its coordinates. In other words, a locational activity recognition system distinguishes between locations only if it helps determine what the user is doing. Typical locations which reflect the user's activity are *workplace*, *home*, *restaurant*, *shopping centre*, etc.

In *motional activity recognition*, the user's activity is abstracted as a motion state or a mode of transportation. Certain motion states can be identified from cellular signals even without any knowledge about the location of observed cell towers. For examples, fluctuations in GSM signals have been shown to be sufficient to determine with reasonable accuracy whether a mobile phone carrier is *walking*, *driving* in a motor car or staying *stationary* [Anderson and Muller 2006a], [Anderson and Muller 2006b]. Accelerometers embedded in recent smartphones can serve the same purpose and further distinguish between finer movements such as *sitting*, *standing* or *running* [Miluzzo et al. 2007], [Miluzzo et al. 2008].

2.2 Inference Techniques

Activity labels are assigned based on *context information* and *background knowledge*. Context information [Dey 2001] includes sensor data, time and user inputs. Background knowledge can either be provided by experts or mined automatically, for example from the Web [Perkowitz et al. 2004]. The set of activities that may be performed at a specific location is an example of useful background knowledge. All activity recognition systems match some context information to activity labels. However, the range of techniques employed during the labelling can vary substantially.

2.2.1 *Where is Inference Performed?.* The architecture of an activity recognition system usually comprises three main components [Choudhury et al. 2008]:

(1) The **sensing module** collects raw sensor data such as audio, video, or communication signals. On mobile phones, GSM/3G and Bluetooth signals can be collected throughout the day.
(2) The **preprocessing module** transforms raw data into a reduced representation called a *feature vector*. The accuracy of the system depends strongly on how data is represented. Features should help differentiate between activities. They can be low-level, for example mean and variance computed on a specific physical signal or high-level, for example the user's abstracted location as estimated from cell tower signals.
(3) The **inference module** takes as input selected features and background knowledge to recognise the user's activity, for example, staying *stationary*, *walking* and *driving*.

Figure 1 shows the three components above in a typical distributed architecture. Sensing is usually limited to the phone's own sensors and preprocessing performed on the phone. Due to the inability of early mobile phones to run complex inference models, this task has traditionally been performed on desktop computers [Eagle and Pentland 2006], [Eagle and Pentland 2009], [Farrahi and Gatica-Perez 2008b], [Farrahi and Gatica-Perez 2008a], [Eagle et al. 2009], [LaMarca et al. 2005]. However, the increasing computational power of mobile phones allowed several recent systems to make inference directly on the phone [Anderson and Muller 2006b], [Anderson and Muller 2006a], [Miluzzo et al. 2007], [Miluzzo et al. 2008]. In both cases, the inference module is the core of an activity recognition system and the place where machine learning models are implemented.

2.2.2 *Early Logic-based Approaches.* Although activity recognition is now considered a research area of its own, it is historically related to plan and goal recognition. Early plan recognition techniques from the late 1970s and early 1980s were consistency-based and made inferences through first-order logical reasoning. In [Schmidt et al. 1978] Schmidt et al. used a simple plan recognition process based on psychological theories of how human observers understand the actions of others. In this process, a single hypothesis is pursued until the observations cannot be matched to expected actions. This triggers a hypothesis revision process. In [Kautz 1987], Kautz et al. developed a formal theory of plan recognition and presented a framework in which the minimum sets of independent plans that entail the observations are inferred. However, the closed-world reasoning they used assumed knowledge of the complete plan library and was unable to adapt to changes in the observed agent's behaviour. Later, Lesh et al. presented a framework based on version space algebra which could recognise novel plans [Lesh and Etzioni 1995]. By repeatedly pruning inconsistent plans and goals when new actions arrived, they also achieved better scalability.

However, none of these logic-based frameworks was used in practice. They indeed had two important limitations. First, they considered all consistent explanations equally and therefore could not determine which one is the most likely. Secondly, they did not handle the noise and uncertainty of real-world data. This is particu-

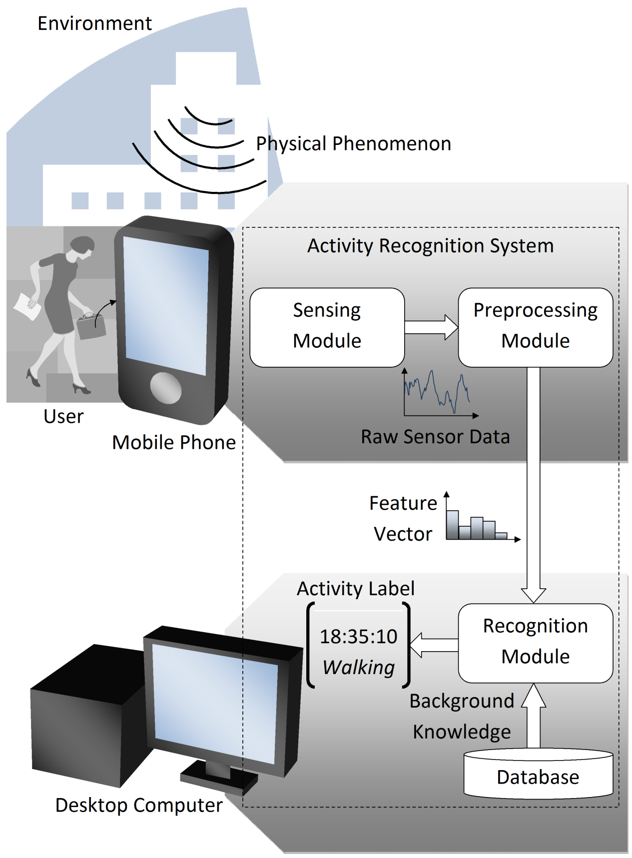

Fig. 1. Typical distributed architectures of a mobile-phone-based activity recognition system. Sensing and preprocessing are performed directly on the phone while inference is made on a destktop computer with greater computational resources.

larly important in our research since there are many sources of noise and uncertainty in mobile phone data. Devices can be turned off, not recharged or forgotten. There are issues with radio communication such as poor indoor reception and fluctuating connections. Cell tower allocation obeys the operator's strategy which is not known a priori. Bluetooth errors include detecting people who are not physically proximate through certain types of windows or doors. There is also a probability that Bluetooth will not discover other proximate devices [Eagle and Pentland 2006]. In order to handle noise and uncertainty, most mobile-phone-based activity recognition systems use statistical models.

2.2.3 *Generative versus Discriminative Models.* Models implemented in inference modules fall in either of two categories. *Generative models* specify a joint probability distribution $P_{X_1 X_2 \ldots X_n Y}$ over features $X_1, X_2, \ldots, X_n$ and the inferred activity Y. They can either model data directly or be used to form a conditional distribution $P_{Y|X_1 X_2 \ldots X_n}$ through the use of Bayes' rule. Model parameters are usually estimated to maximise the likelihood of training data and new instances are classified to the most probable class given the features. Generative models which have been implemented in inference modules include Bayesian Network (BN) [Cho et al. 2007], Hidden Markov Model (HMM) [Eagle and Pentland 2006], [Anderson and Muller 2006b] and Dynamic Bayesian Network (DBN) [Eagle et al. 2009]. In particular, Hidden Markov Models have been learnt directly on a phone using the Baum-Welch Expectation-Maximisation algorithm [Anderson and Muller 2006b].

Discriminative models provide a model only of activities conditional on features. This can either be done by specifying the conditional probability distribution $P_{Y|X_1X_2...X_n}$, or by specifying decision boundaries. Discriminative models which have been implemented in inference modules include Support Vector Machine (SVM) [Farrahi and Gatica-Perez 2008b], Artificial Neural Network (ANN) [Anderson and Muller 2006a] and C4.5 Decision Tree (DT) [Miluzzo et al. 2008] – which have been learnt directly on a mobile phone [Miluzzo et al. 2008].

Although generative models usually require more training data and are sometimes outperformed by discriminative models, they hold a major advantage over the latter in their ability to generate values of any variable in the model. Therefore, generative models can simulate data and provide better understanding of underlying processes.

2.2.4 *Supervised versus Unsupervised Learning.* Models are learnt from training data in essentially two ways. Most activity recognition systems rely on *supervised learning*. In this type of learning, training data is collected by the sensing module and other inputs and transformed into a set of instances by the preprocessing module. Each training instance is then labelled for the user's true activity when data was collected. The supervised inference model generalises from training data to infer the activity labels of unseen instances in test situations. One way to acquire the ground truth is to prompt for activity labels on the fly as the user performs his daily activities [Eagle and Pentland 2006]. This method has the advantage that the information collected is fresh in the user's mind and therefore more accurate. However, providing labels using the interface of a mobile phone is not easy in practice. Also, not all activities can be labelled as they are performed. For instance, interrupting a meeting to input an activity label is not realistic. An alternative approach is to label data *a posteriori*, for example at the end of the day using a diary application or a paper notebook [LaMarca et al. 2005]. The issue with this method are clearly its low temporal accuracy and the limitations of human memory. In either case, activity labels provided by the user are subjective. For example, one user may interpret being at work as *working* while another user may only label as *working* time intervals during which he is actively engaged in his work. Secondly, training an activity recognition system in a supervised fashion requires considerable involvement from the part of the user who not only has to regularly charge his device and carry it with him for several weeks but also needs to spend some time labelling his data.

Unsupervised learning aims to relieve the user from the burden of labelling, usually by clustering data based on some distance measure. One of the most common clustering technique is K-Means, which tries to minimize the total intra-cluster variance [Lloyd 1982]. K-Means takes as a parameter the number of clusters, for example three clusters for *stationary*, *walking* and *driving* [Anderson and Muller 2006b]. However, in many cases, the exact number of clusters depends on the user's personal habits. In particular, this is the case for locational activities. A child may have the three significant locations *home*, *school* and *park* while other users may have many more including *workplace*, *restaurant*, *supermarket* and *library*. More recent algorithms, such as DBSCAN have been used in that context [Zhang et al. 2007]. DBSCAN finds the number of clusters automatically from the estimated density distribution of data points [Ester et al. 1996]. In either case, devising an appropriate distance measure is a difficult problem and can be quite subjective.

3. ACTIVITY RECOGNITION SYSTEMS USING MOBILE PHONES

In this section, we review the design and performance of selected locational and motional activity recognition systems making use of mobile phone data. Each system is categorised according to the criteria stated in the previous section.

3.1 Locational Systems

The first main category of activity recognition systems for mobile phones determines the user's activity in terms of location. We review below the major activity recognition systems for mobile phones which fall into that category. Each system is named after the project in which it was developed followed by a short description.

3.1.1 *Reality Mining: Hidden Markov Model over Cellular and Bluetooth Data.* In the Reality Mining project [Eagle and Pentland 2006], Eagle and Pentland investigate how cell tower information and Bluetooth proximity data can complement each other to help infer important locations and social relationships. The Reality Mining project collected 330,000 hours of continuous data over the course of nine months on 94 mobile phones. Participants, staff and faculty from the MIT Media Laboratory and Sloan Business School, were provided with Nokia 6600 Bluetooth-enabled mobile phones running the ContextLogger. This piece of software was developed by the Context Group at the University of Helsinki [Raento et al. 2005] and records different bits of information including call logs, Bluetooth devices in proximity, the cell tower to which the phone is connected, application usage and phone status, for example, *charging* or *idle*. An anonymised version of the dataset collected during the Reality Mining project is the largest set of mobile phone data publicly available to date.

A significant part of the work carried out in the Reality Mining project falls into the scope of localisation and locational activity recognition. Eagle and Pentland fuse two types of information to determine the user's location. First, relatively high location accuracy is achieved by estimating cell tower probability density functions. When a mobile stays at one place for a long period of time, it connects to different cell towers successively depending on several variables which include signal strength and network traffic. The distribution of detected cell towers varies substantially with changes in location. In order to improve the user's localisation, the authors also incorporate the use of static Bluetooth devices as 'cell towers'. Bluetooth provides spatial accuracy of about 10m and static Bluetooth devices can often be detected in places where cellular signals are weak. For example, Bluetooth desktop computers can be detected in office buildings. Therefore, static Bluetooth devices supplement GSM cell towers in localisation.

In locational activity recognition, Eagle and Pentland consider three activity states – *home*, *work* and *elsewhere* – which correspond to three location clusters in cell tower and Bluetooth data. They build a first-order Hidden Markov Model conditioned on both the hour of the day and the day of the week (weekday or weekend) to capture daily and weekly routines in user behaviour. A Hidden Markov Model [Rabiner 1989] is a generative model in which the system being modelled is assumed to be a Markov process whose state is hidden but output is observed and probabilistically dependent on its state. Hidden Markov Models are widely used in speech recognition [Rabiner 1989] and natural language processing tasks [Leech et al. 1994] to model sequential patterns. However, the first-order Markov assumption also has important limitations for activity recognition. According to this assumption, a state is conditionally independent of all earlier states given the immediately previous state. In other words, a user's activity at a given time slice only depends on his activity at the previous time slice and any longer-range dependency in data is ignored.

The authors report an accuracy typically greater than 95% after training the model on one month of data from several subjects. However, this very high figure is difficult to interpret because neither the exact subset of the dataset used in the evaluation nor the criteria which guided the selection of subjects, training and test periods are detailed. In addition, results are not compared to any baseline. Lastly, the Reality Mining dataset was collected on users from the same academic community. Therefore, generalising results obtained on this dataset is not easy.

3.1.2 *Reality Mining: Principal Component Analysis over Cellular Data.* In a subsequent paper [Eagle and Pentland 2009], the same authors uncover some similarities between the behaviours of different participants from the Reality Mining dataset. Structure in daily human behaviour is represented by the principal components of the complete dataset called *eigenbehaviours*. A user's day is then approximated by a weighted sum of his primary eigenbehaviours. Calculating weights half way through the day of selected users, the authors predict the day's remaining activities with 79% accuracy. In addition, it is shown that users of similar demographics can be clustered into a *behaviour space* spanned by a set of their aggregate behaviours and classified into their correct group affiliation with 96% accuracy – excluding staff members.

Principal Component Analysis (PCA) is an orthogonal linear transformation that performs a coordinate rotation around the data mean to align the transformed axes with the directions of maximum variance called *principal components*. In the new coordinate system, the first coordinate accounts for as much of the variance as possible, the second coordinate accounts for as much of the remaining variance and so on. PCA is commonly used for dimensionality reduction and modelling. In particular, it has been successfully applied to computer vision tasks such as face recognition [Turk and Pentland 1991]. However, PCA has certain limitations for activity recognition. Rare activities, for example, are almost certainly always missed by the first eigenbehaviours and considered noise although they could be the most interesting ones. Also, eigenbehaviours are not robust to variations in the durations of users' activities.

3.1.3 *MULTI: Support Vector Machine over Cellular and Bluetooth Data.* In [Farrahi and Gatica-Perez 2008b], Farrahi and Gatica-Perez consider a subset of 30 users and 121 consecutive days from the Reality Mining dataset to compute both location and proximity features at two different time scales (a fine-grained one every 30 minutes and a coarse-grained one every 3-4 hours). The predictive power of those features is then evaluated in two different tasks. Features are tested alone and in pairs of one location and one proximity feature. Using a Support Vector Machine [Boser et al. 1992] with a Gaussian kernel, the authors aim at (i) classifying a user's day as *weekday* or *weekend* and (ii) classifying a user as a *business* or *engineering student*.

Support Vector Machines perform binary classification by constructing a separating hyperplane which maximises the distance to neighbouring data points from both classes. These neighbouring data points are called *support vectors*. Multiclass classification can be performed by reducing the multiclass problem into several binary problems. Support Vector Machines have been shown to be competitive with many state-of-the-art classifiers notably in computer vision [Osuna et al. 1997] and finance [Huang et al. 2005]. However, training a Support Vector Machine requires to solve a Quadratic Programming (QP) problem which can be computationally demanding. In [Farrahi and Gatica-Perez 2008b], SVMs are not learnt on a mobile phone but on a standard computer.

Evaluation is performed in *leave-one-user-out* cross-validation. Specifically, testing is performed on all the days for one unseen person while training on the data for all other people. Weekdays were best recognised by detecting *work* cell towers whereas weekends were best identified from proximity data (absence of colleagues). Overall, combining location and proximity features yielded to over 80% accuracy, outperforming both location and proximity features alone. Unsurprisingly, group affiliations were best recognised from the identity of people in proximity of the user. The generic locations considered (*home*, *work* and *elsewhere*) were found to be little informative for that purpose. It is likely that considering absolute locations such as a particular building would have been more useful to recognise people's group affiliations but this was not attempted in the experiment. Overall, affiliations were correctly classified with nearly 90% accuracy. However, given that only 6 of the 23 students considered were business students, a most frequent baseline already provides 74% accuracy.

3.1.4 *X-Factor: Dynamic Bayesian Network over Cellular Data.* In [Eagle et al. 2009], Eagle et al. repurpose unsupervised clustering techniques originally developed for community detection to identify salient locations within the network generated by cell tower transitions. Clustering is then validated using data from Bluetooth beacons positioned in the homes of 215 subjects randomly sampled in a U.S. city. Clusters are used as states within several Dynamic Bayesian Networks to predict dwell times within locations and each subject's subsequent movement with over 90% accuracy with an average of 6.88 locations.

A Bayesian Network (BN) [Pearl 1985] is a probabilistic graphical model in which nodes represent variables and arcs encode conditional independences between the variables. The structure of a Bayesian Network can therefore reflect expert knowledge about the classification task. However, static Bayesian Networks do not capture temporal relations in data. A Dynamic Bayesian Network (DBN) [Ghahramani 1998] is a Bayesian Network that models a dynamic system by representing its state at subsequent time slices. Arcs joining nodes at consecutive time slices encode probabilistic temporal relations between them. Dynamic Bayesian Networks can be viewed as a generalisation of Hidden Markov Models and inherit the limitations of the first-order Markov assumption.

In order to detect behaviours that deviate from a given routine, [Eagle et al. 2009] introduces the X-Factor model. This model is really a Dynamic Bayesian Network with a latent variable corresponding to abnormal behaviour. By calculating the entropy of the transition matrix from the X-Factor model, the authors quantify the amount of structure in the daily routines of different demographics and find that there are individuals across demographics who have a wide range of routines in their daily lives. This result runs contrary to previous findings restricted to an academic community [Eagle and Pentland 2006].

The study presented in [Eagle et al. 2009] contrasts with the authors' previous work based on the Reality Mining dataset in that it uses a larger and potentially more representative set of users. Also, the users' locations are clustered in an unsupervised fashion. However, the exact meaning of clustered locations is unknown. A user may stay at a location for a long time or come back to it regularly even if this location has no particular meaning to him for example at traffic lights. The second limitation of the technique is that only very short-range dependencies are considered. Long-range information is only captured in one 'abnormal behaviour' latent variable which switches between two states corresponding to normal or abnormal behaviour. In addition, transition probabilities in the abnormal state are actually smoothed versions of transition probabilities in the normal state. The objective of this approach is to give abnormal behaviour a broader distribution which allows behaviours not seen in training data. However, this approach somewhat gives up in understanding the specificities of abnormal behaviour.

3.2 Motional Systems

The second main category of activity recognition systems for mobile phones outputs the user's activity in terms of motion state. We review below the major activity recognition systems for mobile phones which fall into that category. As previously, each system is named after the project in which it was developed followed by a short description.

3.2.1 *EQUATOR: Artificial Neural Network over Cellular Data.* Certain motion states can be recognised from cellular signals even without knowing where the cell towers that emitted those signals are actually located. Anderson and Muller, from the Mobile and Wearable Computing group at the University of Bristol have developed techniques for the recognition of motional activities using GSM mobile phones. In [Anderson and Muller 2006a], they propose to determine whether a mobile phone carrier is *walking*, *driving* in a motor car or staying *stationary*. In order to allow their system to run on any mobile phone, the authors avoid the use of accelerometers. Instead, they rely on two simple observations. First, in a given environment, the rate of cell change increases with speed. Secondly, the level of signal strength fluctuation at the same physical position and the variance of signal strength levels at different physical positions are greater when the phone carrier is moving compared to when the phone is left stationary.

The method presented in [Anderson and Muller 2006a] is built around an Artificial Neural Network (ANN) [Jain et al. 1996]. Artificial Neural Networks are non-linear statistical models which simulate the function of a biological neural network. Due to their ability to model complex relationships between inputs and outputs, ANNs have been applied to many pattern recognition problems including face detection [Rowley et al. 1998] and handwritten text recognition [Garris et al. 1998]. However, the architecture of an ANN makes the integration of background knowledge and the understanding of a learnt ANN arduous in general. The structure of the ANN used in [Anderson and Muller 2006a] comprises an input layer, an output layer and a single hidden layer of eight units. Following the authors' observations, the ANN is fed with two features, computed over a time interval ranging from 15 to 30 seconds: (i) the number of distinct cells monitored over the time interval and (ii) the sum of signal strength fluctuation across currently serving and neighbouring cells. Given these two input features, the network outputs the user's current activity. The authors observe that, at times, the patterns of fluctuation while *driving* match those observed while *walking*. This happens in urban areas because of traffic lights and traffic congestion which force vehicles to slow down and stop. The effect of such events on the performance of the classifier is mitigated by applying a simple averaging filter to the input.

Anderson and Muller train their ANN in a supervised fashion, by repeatedly presenting the system with data corresponding to each activity. The weights of the ANN are then learnt using back-propagation, a gradient-based method in which errors in classification of training data are propagated backwards through the network to adjust the bias weights of the network elements until the mean squared error is minimized. Due to the computational cost of the algorithm, the authors were not able to train the ANN directly on a mobile phones and had to use a desktop computer.

One of the shortcomings of the technique is that the ANN must be retrained in every type of environment. Indeed, network capacity in dense urban environments is increased by using a higher number of short-range cells. Therefore, the same rate of cell change and level of signal fluctuation can correspond to different activities dense and sparse urban environments. In the absence of a scheme to share data across users, collecting and labelling training data while *walking*, *driving* and staying *stationary* in each visited place could rapidly become intractable.

Also, in real life, certain sequences of motion states are much more likely than others. For example, a short walk followed by a long drive followed by another short walk occurs almost every time a mobile phone carrier uses a car. In contrast, a succession of numerous short drives interrupted by short walks is very uncommon. Since the approach does not take into account temporal patterns of activities, any sequence of activities can be output, no matter how inconsistent or unlikely it is.

3.2.2 *EQUATOR: Hidden Markov Model over Cellular Data.* In a subsequent paper [Anderson and Muller 2006b], Anderson and Muller present an unsupervised activity recognition technique in which the user is still required to spend about 15 minutes *walking*, *driving* and *stationary* in each environment but does not have to keep a record of his activities anymore. As a calibration step, K-means [Lloyd 1982] is applied to partition observations into three clusters according to cell and signal strength fluctuations. The three means of these clusters are assigned to the three activities using a fixed ordering defined in background knowledge. Each observation is then labelled with the activity of its cluster and a five-step Hidden Markov model is learnt over this data using the Baum-Welch algorithm [Baum et al. 1970]. After training the HMM, the most likely state sequence producing a sequence of observations is determined using the Viterbi algorithm [Viterbi 1967]. In comparison with the authors' previous approach, the use of an HMM gives an important advantage in that it models not just a single activity but temporal sequence of activities such as *driving* followed by *walking*.

Both this method and the previous were implemented on an Orange SPV C500 mobile phone capable of monitoring up to six neighbouring cells in addition to the current serving cell. This phone model has a 200Mhz ARM OMAP 730 processor. Unlike the ANN used in the previous method, the HMM could be learnt directly on the phone. K-means and Baum-Welch were both implemented on the phone and training took approximately two minutes to complete. Both the trained ANN and HMM recognise activities on the phone in real time.

The performance of both techniques was significantly higher in urban than in metropolitan areas. An explanation for this is that cell and signal strength fluctuations are much higher within metropolitan areas, due to high concentrations of micro-cells and high traffic. As a result, a number of *stationary* states were misclassified as *walking*. For both techniques and in both environments, the poorest results were obtained for the *driving* activity. In towns, cars often have to slow down. As a result, it is sometimes difficult to distinguish between driving and walking fast.

Overall, the unsupervised method achieved the same level of performance as the ANN without the need of manually labelled data. This is a key advantage for mobile users who cannot label their data easily. However, the unsupervised calibration process relies on a hard-coded ordering of the means associated to each cluster. If, as the authors suggest, more activities were recognised, including, for example, *running* and *cycling*, signal strength fluctuations associated to certain activities may be much closer and difficult to order consistently over all observations.

Also, the extension of the strategy to 3G networks is an open issue. 3G networks use smaller cells because they have to support the transmission of large amounts of information and operate at a higher frequency. In that respect, the decrease in performance observed when the density of GSM cells increases raises some concerns. 3G cells also expand and contract in size depending on the number of simultaneous calls being made. Signal strength fluctuations will therefore not always be indicative of motion.

Anderson and Muller showed three interesting things. First, motion information could be captured on a GSM mobile phone without the need of additional sensors. Secondly, the burden of data labelling could be avoided by using unsupervised calibration. Lastly, HMMs could be both learnt and used for inference on a mobile phone. Nevertheless, in comparison to accelerometer-based techniques, the methods they propose are significantly slower. This is due to the length of the time interval over which features are computed (15 to 30 seconds). As more and more mobile phones embed accelerometers, Bluetooth and GPS, using GSM data only may be an unnecessary constraint. Patterson et al. [Patterson et al. 2004], for example, have been able to distinguish between a wider range of transportation modes using a GPS receiver.

3.2.3 *Place Lab: Boosted Logistic Regression with Decision Stumps over Cellular Data.* Place Lab is software developed by Intel Research Seattle and academic research partners for device positioning. Sohn et al., at the University of San Diego have contributed to Place Lab and proposed a mobility detection method based on GSM traces [Sohn et al. 2006]. Similarly to Anderson and Muller, they propose to recognise if the carrier of an unmodified mobile phone is *walking*, *driving*, or staying *stationary*. The recognition scheme and features used in this work are more elaborate than previous approaches and achieve an overall accuracy of 85%. However, these techniques have not actually been implemented on a mobile phone.

In [Sohn et al. 2006], activity recognition is based on the same principle as fingerprint-based location estimation. Specifically, radio signals observed from fixed sources are consistent in time, but variable in space. Given a series of GSM observations, a change in the set of visible cell towers and signal strengths is therefore interpreted as motion. A Euclidean distance is designed to capture the similarity between consecutive GSM measurements and the authors find a proportional relationship between this distance and speed of movement. Seven statistical measures are computed as features. Three of them compare two consecutive measurements, while the other four use a sliding window of measurements varying between 10 and 300 seconds. Averages computed on larger windows are more robust to noise. However, they are also less affected by speed variations. The window size is therefore a trade-off between robustness to noise and responsiveness.

The seven features considered feed a two-stage classification scheme. A first classifier categorises an instance as *stationary* or not. If the instance was not classified as *stationary*, a second classifier determines if the user was *walking* or *driving*. Both classifiers implement Boosted Logistic Regression (LR) with Decision Stumps [Friedman et al. 2000] as the weak classifiers. Boosting is a general method to construct an accurate classifier by combining the output of so-called *weak classifiers* and has been used successfully in face detection [Viola and Jones 2004]. In [Sohn et al. 2006], weak classifiers are depth-one Decision Trees, equivalent to if-then-else decision rules. The authors used the Weka machine learning toolkit [Witten and Frank 1999] to compare their method with several machine learning classifiers and selected the one technique which achieved the highest accuracy.

GSM data was collected on Audiovox SMT mobile phones running a custom logging application. Each reading included identifiers, signal strength values and channel numbers of up to seven nearby cell towers. In addition, the application recorded the channel numbers and associated signal strength values of up to 15 additional channels. Data collectors were three members of a research team who went about their daily lives for one month. The selection of these data collectors is justified by the very tedious and error-prone self-reporting protocol. However, the fact that an external user could not train the supervised algorithm raises some doubts on the possibility to deploy such a system in the real world. Each user had to report his walks, drives, and stays using a custom diary application and a paper notebook.

The method achieved very high precision and recall when detecting *stationary* activities (95.4% and 92.5%, respectively). *Walking* and *driving* were also quite well recognised with recalls of 80% and 81.7%, respectively. However, the precision figure for *walking* (70.2%) was much lower than for other activities. As in [Anderson and Muller 2006b] and probably for the same reasons, many walks were misclassified as *driving* activities. It can be observed that motional activity recognition systems based on GSM signals are never evaluated in rural areas. Methods based on cell and signal strength fluctuations may be challenged in such regions which are covered by macro-cells with ranges of up to several kilometres.

Building on their activity recognition method, the authors also design a step counter. The application is based on a simple heuristic obtained by performing linear regression with a 5-fold cross-validation on their dataset. The accuracy of the step count method was evaluated against the measurements of the Omron Healthcare HJ-112, a highly-rated commercial pedometer. Overall, 50 days of data were collected. The heuristic predicted daily step counts ranging from 1,500 to 12,000 steps, with an average of 5,000 steps. Comparing with the pedometer's step counts, an average difference of 1,400 steps per day (with a standard deviation of 900 steps) was observed. Therefore, a specific effort from the user is very likely to go unnoticed (the average error is equivalent to almost an hour of walk). Also, the correlations between measured and predicted step counts for the three data collectors were relatively low.

3.2.4 *CenceMe: Supervised Decision Tree over Accelerometer Data.* The CenceMe system [Miluzzo et al. 2007], [Miluzzo et al. 2008] combines inference of activity, travel, conversation and presence of individuals using a Nokia N95 mobile phone. This information can then be shared through social networking application such as Facebook [Facebook, Inc. 2009] and MySpace [MySpace, Inc. 2009]. CenceMe implements a split-level classification scheme whereby inference is run in part on the phone and in part on a backend server to improve scalability. Activity recognition is performed directly on the phone using on-board accelerometer data to determine whether the user is *sitting*, *standing*, *walking* or *running*. The accelerometer sensor and event detector are Symbian C++ modules that act as daemons producing data for corresponding JME client methods. A preprocessing module fetches raw accelerometer data from the local storage component and extracts lightweight features including the mean, standard deviation and number of peaks of the accelerometer readings along the three axes of the accelerometer. CenceMe's inference module is based on a C4.5 Decision Tree [Quinlan 1993] which is trained off-line on a desktop machine because of computational requirements. Decision Trees are tree-like graphs in which each node stands for a test on the value of a feature and each branch represents a possible outcome of a test. The class of an instance can be read by following branches from the root of the decision tree down to one of its leaves. Decision Trees are easy to interpret and quick heuristics exist to build them. However, Decision-Tree learning can produce large trees which overfit training data and do not generalise well. In practice, pruning techniques are used to avoid overfitting. In [Miluzzo et al. 2008], the output of the training algorithm is a small depth-three tree which classifies test instances in less than a second on average on the phone.

CenceMe was evaluated in a small-scale supervised experiment involving eight users, student and faculty from Dartmouth College. These users annotated their actions over a one week period at intervals of approximately 15 to 30 minutes. With an average accuracy of 78.89%, reported figures are up to 20% lower than those reported using custom hardware [Lester et al. 2006]. In particular, the system has difficulties differentiating between *sitting* and *standing* and *walking* and *running*. The position of the phone was found to impact recognition accuracy. Specifically, holding the phone in a trouser's pocket or at the belt produces similar results but having it at a lanyard position yields poor accuracy when classifying *sitting* and a slightly lower accuracy for *running*. The length of the lanyard cord and its type were also found to affect the results.

The CenceMe system is one of the existing implementations which actually runs on a mobile phone. However, the techniques implemented are relatively rudimentary and therefore achieve low accuracy even on a custom small-scale controlled experiment.

3.3 Summary Table

For each system we reviewed, Table I summarises sensor inputs, inference technique, outputs and evaluation details. It is striking to observe that activity recognition systems for mobile phones can achieve relatively high accuracy but only address a small set of human activities. Consequently, many of the applications mentioned in the introduction section cannot be implemented properly. For example, an automatic diary unable to recognise meetings would be of little use in an office environment. In most cases, locations and motion states act as substitutes for activities. The problem for potential applications is that these approximations can be quite inaccurate. For example, *staying at home* does not tell us much about what the user is doing and *being at work* is not equivalent to *working*. It is often believed that finer activity recognition will be achieved through augmented sensing capabilities. However, one cannot but notice that certain sensors which have been supported by standard mobile phones for several years, are still very little used. In particular, this is the case of the Bluetooth sensor. In theory, Bluetooth short-range communication could provide very useful information for activity recognition including the co-presence of mobile Bluetooth devices. For instance, the co-presence in a meeting room of phones belonging to certain team members indicates that a meeting is taking place. In the next section, we examine what other sensing technologies may be used for mobile-phone-based activity recognition in the future.

4. PROSPECTS

Each new generation of mobile phones brings advances in both hardware and software. Looking at trends in high-end models gives an insight into what may be available for activity recognition on every mobile phone in the near future. In this section, we review some work carried out in activity recognition using technologies which are expected to be widely available on standard mobile phones in the short term.

		Locational Systems				Motional Systems			
		Reality Mining [Eagle and Pentland 2006]	Reality Mining [Eagle and Pentland 2009]	MULTI [Farrahi and Gatica-Perez 2008b]	X-Factor [Eagle et al. 2009]	EQUATOR [Anderson and Muller 2006a]	EQUATOR [Anderson and Muller 2006b]	PlaceLab [Sohn et al. 2006]	CenceMe [Miluzzo et al. 2008]
Sensing	Mobile phone	N6600	N6600	N6600	N/A	SPV C500	SPV C500	SMT 5600	N95
Sensing	Cellular signals	X	X	X	X	X	X	X	
Sensing	Bluetooth signals	X		X					
Sensing	Accelerometer								X
Inference	Model	HMM	PCA	SVM	Ncuts DBN	ANN	K-means HMM	Boosted LR	C4.5 DT
Inference	Unsupervised				X		X		
Inference	Trained on phone						X		
Inference	Tested on phone					X	X		X
Output		Home Work Elsewhere	Home Work Elsewhere	Weekend Weekday	7 clusters	Stationary Walking Driving	Stationary Walking Driving	Stationary Walking Driving	Sitting Standing Walking Running
Evaluation	Public dataset	X	X	X					
Evaluation	Time period	1 month	N/A	121 days	1 month	9 hrs	16.5 hrs	1 month	1 week
Evaluation	Number of users	N/A	N/A	30	215	N/A	N/A	3	8
Evaluation	Percent accuracy	95	79	80	90	89	80	85	79

Table I. Comparison of mobile-phone-based activity recognition techniques.

4.1 Extended Sensing Capabilities

High-end mobile phones are equipped with an increased number of sensors. Many of those sensors have been used for activity recognition on non-mobile-phone-based systems. We give below some examples of how the GPS, accelerometers and RFID have helped in activity recognition and discuss their added-value in mobile-phone-based activity recognition.

4.1.1 *Accurate Outdoor Localisation and the Global Positioning System.* The Global Positioning System (GPS) is a global navigation system relying on a constellation of satellites around the Earth that transmit radiowave signals. GPS positioning has a precision of about 15 metres on the surface of the earth and is widely used for navigation purposes. Anticipating a boom in localisation services, smart phone manufacturers have equipped an increasing number of high-end models with GPS receivers. However, few activity recognition systems for mobile phones make use of GPS information at the present stage. One notable exception is AniDiary [Cho et al. 2007] which summarises a user's day in a cartoon-style diary based on GPS localisation mapped to the nearest building. GPS has also been utilised independently of any mobile phone. Assuming the availability of a street map, Patterson et al. [Patterson et al. 2004] propose to predict both a person's location and mode of transportation to help guide cognitively impaired people safely using a portable GPS unit. In [Liao et al. 2007], Liao et al. segment a GPS trace in order to generate a discrete sequence of activity nodes. Specifically, a person's activity is labelled everytime he passes through or stays at a 10 metre patch of the environment. The technique they present is capable of detecting both motional activities such as *walking*, *driving car* and *riding bus* and locational activities including *work*, *leisure*, *sleep*, *visit*, *drop off/pickup*.

In semi-urban areas, GPS provides an outdoor position with a precision of about 20 metres which allows the systems mentioned above to recognise more varied locational activities than GSM-based systems. For example, AniDiary attempts to infer events such as *doing sport*, *eating out* and *shopping*. The higher accuracy of GPS in outdoor environments makes it possible to delimit finer-grained locations including shops, hospitals or schools [Cho et al. 2007]. Some of those fine-grained locations such as *parking lots*, *bus stops*, *roads* and *bus lines* help in turn determine fine-grained motional activities including whether the user is *in a car* or *riding a bus* [Patterson et al. 2004], [Liao et al. 2007].

Unfortunately, GPS has been shown to be available for less than 5% of a typical user's day [Sohn et al. 2006]. Also, GPS signal is affected by a number of factors including atmospheric conditions and the local built environment. In particular, GPS signals reflect off buildings in cities in what is known as the multipath effect. As a result, the accuracy of activity recognition can be very low. AniDiary, for example, was tested on 27 days of real data from one student equipped with a Nokia Series 60 smart phone running the ContextPhone application [Raento et al. 2005]. Its correct detection ratio was only 34.1%. In [Patterson et al. 2004], vehicular journeys are much better recognised than pedestrian ones. The multipath effect is indeed known to be easier to eliminate when the user is in a moving vehicle.

Some applications such as Google Latitude [Google, Inc. 2009b] illustrate the high interest that service providers have in the GPS technology. However, the use of GPS in activity recognition is conditioned on the availability of high-capacity batteries. The 1Hz update rate used in [Liao et al. 2007] could indeed easily drain the battery of a mobile phone in a few hours. In addition, GPS must be supplemented by other technologies in indoor environments. Lastly, the cost of GPS receivers is still high and GPS-enabled mobile phones are therefore unlikely to be common around the globe in the very short term. With multilateration, GSM positioning achieves relatively high precision at no supplementary cost and already has billions of potential users.

4.1.2 *Recognising Body Postures and Motions Using Accelerometers.* Beside GPS, another technology could well bring new advances in motional activity recognition. Many high-end smart phones are now equipped with accelerometers which measure the phone acceleration in several directions to allow more natural forms of interaction with the user. Moving the phone upwards or downwards can for example scroll the screen on the Apple iPhone. Accelerometers help recognise body movements, such as *standing up*, *sitting down*, *laying down*, *climbing stairs* or *taking a* *lift*, which cannot be achieved using cellular signals alone. CenceMe, for example, recognises when the user is *sitting*, *standing*, *walking* and *running* [Miluzzo et al. 2007], [Miluzzo et al. 2008]. Accelerometers have also been used for activity recognition on body-sensors [Ward et al. 2006] and custom activity recognition platforms such as the Mobile Sensing Platform (MSP) [Choudhury et al. 2008]. However, wearing body-sensors can be inconvenient for the user and custom sensing platforms represent an additional investment and device to carry.

The most important technical challenge with accelerometers is their sensitivity to placement. Body-worn accelerometers were shown to achieve higher posture recognition when worn at the hip [Bao and Intille 2004] whereas a mobile phone is often left in a pocket, in a bag or on a desk. Therefore, CenceMe has difficulties distinguishing between *sitting* and *standing* or *walking* and *running* when the phone is held in a trouser's pocket, at the belt or at a lanyard position. In [Reddy et al. 2008], augmenting an N95's accelerometer data with GPS information allowed to determine whether the user is *stationary*, *walking*, *running*, *biking*, or in *motorised transport* without strict position/orientation requirements. However, this work was based on only twenty hours of data from six users.

Processing accelerometer data induces a significant computational overhead. In order to make their system work on mobile phones, the designers of CenceMe used lightweight features and a very simple classification technique which performs 20% lower than state-of-the-art accelerometer-based behaviour inference [Lester et al. 2006].

Accelerometer-based activity recognition is primarily motivated by healthcare and assembly applications. In both cases, wearing special equipment is a reasonable assumption. For everyday activity recognition using mobile phones, placement seems to be an important constraint which strongly affects recognition accuracy. In addition to uncertainty relative to the user's activity, mobile phone systems would have to handle uncertainty about the placement of the phone.

4.1.3 *Activity Recognition through Object Uses and Radio-frequency Identification.* Radio-frequency Identification (RFID) is a technology which allows the detection of an object using radio waves. Integrated circuits called RFID tags are incorporated into objects in order to make them detectable by an RFID reader. There are two types of tags. Active tags embed a battery and can be read from several metres away while passive tags are inexpensive, have no battery and a shorter range of about 10 centimetres. Commercial applications of the RFID technology have been in the field of logistics and include the detection and tracking of products, animals or people.

RFID has recently been used to recognise activities from object uses in smart homes [Wu et al. 2007], [Patterson et al. 2005]. Detecting that the user has been close to certain ingredients and kitchenware, for example, is an indication that he is currently cooking. The sequence in which objects are used is also useful information which allows the distinction between close activities such as preparing different meals. No other existing technology allows such a fine granularity in activity recognition.

Certain recent phone models such as the Nokia NFC (Near Field Communication) embed an RFID-compatible reader and tag which allows its users to make contactless payments, share media, read tagged posters or use their mobile phones as tickets. Although no activity recognition system for mobile phones makes use of the RFID technology at the moment, it is very likely that the new opportunities offered by RFID-enabled phones will be exploited in the near future.

As in the case of accelerometers, the placement of RFID readers may be a problem on mobile phones. Passive tags only have a very short range of a few centimetres. Therefore, detecting objects used when the phone in a trouser's pocket is very unlikely. The recent availability of watch phones could however make embedded RFID readers much closer to objects held in hand.

4.2 Multimedia Activity Recognition

Mobile phones have become multimedia platforms capable of recording audio and video streams. Many mobile phones are already equipped with one or two cameras and the number of cameraphones is expected to boom in coming years [Mawston 2008]. Cameras and microphones have both been shown to help identify a person's activity. However, for a variety of reasons, few mobile-phone-based activity recognition systems take advantage of audio and video at the moment. In this section, we give an overview of what is currently achieved in video and audio-based activity recognition, analyse why these techniques have not been used so far together with mobile phones and anticipate future applications.

4.2.1 *Vision-based Activity Recognition Using Cameraphones.* The literature in vision-based activity recognition is so rich that vision-based activity recognition can be considered a domain of its own. Vision-based systems apply computer vision techniques to detect human bodies and their activities usually on videos from standard surveillance cameras [Robertson and Reid 2006], [Niu et al. 2004], stereo-cameras [Harville and Li 2004] or infra-red cameras [Han and Bhanu 2005]. Many vision-based systems such as [Park and Kautz 2008] focus on Activities of Daily Living which have been shown to be useful in a number of healthcare situations. One other typical application is to detect falls [Nait-Charif and McKenna 2004], [Williams et al. 2007] in home settings. A third set of vision-based systems address office situations and recognise activities in meetings [Wallhoff et al. 2004], [McCowan et al. 2005]. Lastly, a large part of vision-based systems attempt to track pedestrians and detect abnormal or suspicious behaviours, for example, on critical sites [Bodor et al. 2003].

Although many of the phones sold each year around the world are now camera phones, cameras are still little used by activity recognition systems for mobile phones. There are several reasons for this. First, unlike surveillance cameras, embedded cameras are not turned on continuously to preserve the battery of the phone. Therefore, images available for inference are either captured occasionally by the user [Cho et al. 2007] or can be captured at regular intervals [Miluzzo et al. 2008]. In any case, most visual inputs are fixed images, not video streams. In practice this makes computer vision problems such as body or face detection much harder because many false alarms cannot be filtered out. Secondly, certain image processing tasks, including segmentation, body detection and tracking are computationally expensive. Running them in the background of a mobile phone together with another application in the foreground is challenging. Lastly, images captured by camera phones are unconstrained. This is a very important difference with surveillance cameras which are usually fixed or have a predefined movement. On such cameras, the background can be subtracted to identify new elements in the image. This cannot be performed on pictures taken anywhere, at any time and in varying lighting conditions.

Integrating visual information into activity recognition on mobile phones is a challenge which has not been seriously taken up so far. The only use of visual information has been limited to reading the metadata of a photograph [Cho et al. 2007] and computing the overall brightness of an image captured from a mobile phone using a backend processing tool on a desktop computer [Miluzzo et al. 2007], [Miluzzo et al. 2008]. With a large part of mobile phones embedding one or two cameras and featuring high computational capabilities, a lot more could be achieved. However, capturing images automatically throughout the day raises important privacy issues. In addition, many images would be of little use, for example, when the phone is in a pocket or in a bag. Opportunistic strategies could be implemented to take advantage of any image or video captured by the user himself. Alternatively, the time interval between consecutive snapshots could vary dynamically depending on context. A better idea may however be to integrate into the activity recognition process visual information captured by other cameras such as fixed surveillance cameras or webcams.

4.2.2 *Listening to Activities through the Mobile Phone Microphone.* One sensor which has been present in mobile phones since their inception is the microphone. The amount of information about a person's activity that can be inferred by a human listener from a conversation or an audio record is considerable. Bugging has therefore become common in espionage and police investigations. With its communication capabilities and built-in microphone, the mobile phone has all the properties of a covert listening device. Surprisingly, built-in microphones have been very little used in activity recognition. CenceMe [Miluzzo et al. 2008] uses Matlab processing on the back end to generate a noise index from audio samples captured on N80 and N95's microphones. Using a voice detection algorithm on audio streams captured from a phone in the user's pocket in custom experiments, the system correctly detects if the user is engaged in a conversation with an accuracy of almost 80%. Other examples of mobile microphones used for activity recognition include the MSP [Choudhury et al. 2008] and body-worn microphones [Ward et al. 2006] in industrial environments where certain activities such *sawing*, *hammering*, or *drilling* have a distinctive audio fingerprint. However, low recall and precision rates of 66% and 63%, respectively, were obtained.

Capturing and processing audio faces similar privacy and technical issues as processing video streams. Capturing audio through the phone's built-in microphone is battery-consuming and processing the input to detect speech or activity fingerprints involves heavy signal processing. However, unlike cameras, microphones do not necessarily need to be oriented accurately and their placement seems to be less critical for certain tasks such as conversation detection [Miluzzo et al. 2008]. The main obstacles to activity listening are therefore the battery capacity and processing capabilities of mobile phones, which will inevitably increase in the coming years. It can therefore be expected that built-in microphones will be used for activity recognition in the short term with great benefits.

4.3 Discussion

As more and more sensing technologies are integrated into high-end mobile phones, the question of their use in activity recognition cannot be avoided. GPS, accelerometers and RFID used independently of any mobile phone, have been shown to allow the distinction of activities which have not been distinguished using cellular, Bluetooth or Wi-Fi signals alone. It can be expected that in the coming years, fine-grained locational and motional activities will be distinguished using GPS and accelerometers, respectively. The case of RFID is somewhat different since integrated RFID readers are constrained by the fact that mobile phones are most often carried in a pocket or handbag and therefore out of RFID range for passive tags incorporated in everyday objects. In addition, the use of RFID is actually conditioned on the wide integration of RFID tags in everyday objects, which is still to happen.

In order to go beyond locational and motional activity recognition, multimedia technologies already present in today's phones seem to be the best way. Today's smart phones feature computational capabilities equivalent to that of standard desktop machine only ten years ago. It therefore becomes possible to process video and sound signals directly on the device. The major limitations then become battery power and privacy issues. The increase in capacity of mobile phone batteries in recent years indicates that it is only a matter of time before continuous audio streams can be recorded and processed throughout the day. Audio processing could provide not only voice and event detection but also speaker and speech recognition, opening the way to text mining on transcriptions – even if those transcriptions are incomplete or noisy. The Google Voice application for mobile phones already allows the transcription of voicemail into text [Google, Inc. 2009c]. Supplementing location and co-presence with information about the content of conversations would allow the recognition of new activities such *making a presentation*, *introducing somebody*, *chatting* about a particular topic, or even *saying a prayer*, with obvious privacy implications.

Existing activity recognition systems for mobile phones only capture information from embedded sensors. For certain sensors, such as cameras, the built-in placement may not be optimal. Integrating external sensors such as webcams or computer microphones into activity recognition through network connectivity may need to be considered. Such an approach actually better corresponds to the original visions of ubiquitous computing [Weiser 1991] and would be a sign of the shift from mobile to ubiquitous computing.

In any case, the integration of additional sensors has a cost in terms of battery, memory and computational resources. It is therefore necessary to balance this cost with expected benefits for activity recognition. In certain cases, the information provided by two sensors can be redundant. GSM localisation for example can achieve high accuracy in cities through multilateration without the need of GPS. In other cases, sensor readings can be complementary. For example, multimodal sensing has been shown to offset the information lost when sensor readings are collected from a single location [Maurer et al. 2006], [Lester et al. 2006]. Also, combining several location-sensing modalities increases coverage [Papliatseyeu and

Ibarra 2008]. Lastly, considering two cues can help differentiate between more activities. Combining accelerometer data and GSM localisation can for instance help differentiate between several *sitting* activities such as *reading at the library*, *working at the office computer*, or *having lunch at the restaurant*. Determining the subset of sensors which are most useful and least redundant for the recognition of a given set of activities is an important question which has not been methodically addressed yet.

5. RESEARCH CHALLENGES

Given the extended sensing and computational capabilities of today's smart phones and the potential benefits of activity recognition, one could legitimately ask why activity recognition systems have not made it into the mainstream market yet. There are of course certain privacy issues associated to following people's behaviour continuously through their mobile phones. However, it is striking to observe that *location-based services*, which face similar privacy issues, are gaining considerable momentum at the same time. In this section, we discuss obstacles to the development of mobile-phone-based activity recognition and make out some directions for future research.

5.1 Handling Missing Data

There are many reasons for a mobile phone to stop collecting data. Some of these reasons are due to fundamental constraints of mobile phones such as limited battery life. GSM signals may also be unavailable in certain indoor locations. Lastly, the user may decide to shut down his device, for example, in a theatre. In all these cases, whether logs are kept locally on the phone or remotely by the operator makes no difference. Local logging also has to face specific problems such as memory consumption, unstable operating systems and conflicts with other applications which may cause the logging daemon to crash or malfunction. In order to recognise the user's activity over long intervals of missing data, one has to make use of long-range information.

5.2 Exploiting Long-range Dependencies

Looking several hours back in time can often help recognise a user's activity [Choujaa and Dulay 2009]. For example, the user commuting to work in the morning is a strong indication that he will commute back home in the evening, whatever he may do in the meantime. Similarly, if the user came back home late at night, he is unlikely to wake up and go to work early the next morning.

This type of common-sense reasoning performed by early logic-based plan inference systems is paradoxically out-of-reach for current statistical activity recognition models which ignore any long-range relationship between activities. For example, most probabilistic activity recognition models only consider dependencies between variables at a small number of subsequent time slices, typically one or two. In particular, this is the case of Hidden Markov Models and Dynamic Bayesian Networks designed for activity recognition. By ignoring longer-range dependencies, research in activity recognition misses out one of the specific advantages of the mobile phone platform, that is, its ability to record sensor data continuously throughout the user's day.

5.3 Exploiting Interpersonal Dependencies

There are many reasons why the behaviours of individuals belonging to the same group or community should be related. Students attending the same classes have to be at the same lecture halls at the same times. Friends having lunch together meet at agreed times on a regular basis and go to the same restaurant. Work meetings gather work colleagues at a given place for the same duration. In [Pentland 2007], affiliations and 'close' friendships in the Reality Mining dataset were accurately identified from the conditional probability of a participant's activity given other participants' activities at the previous time slice. Combined with proximity data, the conditional probability structure of location data was also shown to help predict whether two conference attendees were affiliated with the same company [Gips and Pentland 2006]. However, the extent to which a person's activity can be predicted from related people's activities throughout the day and the week requires further research.

5.4 Overcoming Temporal Variations in Human Behaviour

The time spent on activities varies from one day to the next, even if those days follow the same daily pattern. For example, a user may stay longer at work on a particular working day because of extra work or technical difficulties in the public transport system. Similarly, a user may spend longer at the restaurant if he has a guest. The reasons behind temporal variations are diverse and not necessarily detectable in mobile phone data.

Temporal variations make the comparison of human behaviours more difficult. Simple approaches such as comparing the user's activity at fixed times on different days do not work well. The standard approach to accommodate for temporal variations in human behaviour is to consider coarse-grained time slots. In [Farrahi and Gatica-Perez 2008a], for example, the following eight coarse-grained time slots are introduced: 0-7am, 7-9am, 9-11am, 11am-2pm, 2-5pm, 5-7pm, 7-9pm, and 9-12pm. These coarse-grained time slots do remove some minor temporal differences between daily routines but this is achieved at the price of a much lower temporal resolution and the boundaries of these time slots are arbitrary. Accommodating for temporal variations without sacrificing temporal accuracy is an open problem.

5.5 Evaluating Activity Recognition in Applications

The present basis for the evaluation of activity recognition systems is the *exact match*. This criterion is appealingly simple but suffers some serious drawbacks. For example, let us consider an activity recognition system that distinguishes between staying *stationary*, *walking* and *running*. Outputting *walking* instead of *running* is considered equally erroneous as outputting staying *stationary*. However, if the final application is to estimate the user's physical expenditure, the first error is clearly less critical than the second.

Activity recognition is most often an intermediate task contributing to a larger application. Therefore, two kinds of evaluations can be carried out. First, one may assess the contribution of activity recognition to a specific application. This approach is particularly suited when the contribution of activities to the final application is indirect or difficult to perceive, for example, when the user's activity is an item of context among others. Secondly, activity recognition systems may be tested independently of any application on testbeds specifically designed for this purpose. Most mobile-phone-based activity recognition systems are tested using the second approach which allows deeper understanding of difficulties and guarantees that an activity recognition technique can be employed in diverse situations. However, it also demands a certain consensus on test activities.

5.6 Establishing a Standard Set of Activities

Mobile-phone-based activity recognition is not yet standardised. Many studies in activity recognition using body sensors and video cameras focus on a subset of the well-defined Activities of Daily Living (ADL) [Katz et al. 1963], [Katz 1983] which have been shown to help in healthcare, for example [enotes.com, Inc. 2009]. No consensual set of activities has yet emerged in mobile-phone-based activity recognition. There are at least two reasons to this. First, mobile phone hardware evolves so rapidly that authors cannot predict what will be achievable medium term. For example, a whole new range of opportunities will open up if standard mobile phones are equipped with accelerometers. Secondly, potential applications for mobile-phone-based activity recognition are very diverse. Therefore, it is difficult to define a single set of activities that would be suitable for all of them. Nevertheless, it seems that identifying some useful activities which can be recognised using today's mobile phones would have the benefit of setting an objective and guiding research efforts.

5.7 Collecting Reference Testbeds

Research in mobile-phone-based activity recognition is short of publicly-available datasets and reference testbeds. Gathering real-life context and activity data faces four major problems. First, designing an application which needs to run robustly for long hours on a mobile phone demands specific programming skills. Secondly, technical difficulties may arise both in hardware and software during data collection. Thirdly, logging activities requires considerable user involvement and reporting. Typically, users forget to charge or carry their devices and report sparse or erroneous information [Eagle and Pentland 2006]. Lastly, some pieces of context information are personal. Considering potential misuses of personal data, logging an individual's successive locations, contacts and activities raises legitimate concerns.

Privacy issues are often used as an argument against the public release of an existing dataset. This situation prevents a lot of experiments from being repeated. As a result, it is actually difficult to tell whether mobile-phone-based activity recognition is actually making any progress. Publishing anonymised versions of existing datasets would be a driver for reasearch in mobile-phone-based activity recognition and allow existing systems to be compared on the same testbed. In that respect, the CRAWDAD initiative [Yeo et al. 2006] for location data is exemplary.

6. CONCLUSION

As a research field, mobile-phone-based activity recognition is still relatively immature. In particular, the absence of reference testbeds and the vagueness of its objective are revealing. However, we believe that the task could be instrumental in the realisation of the ubiquitous computing vision. First, the mobile phone is one of the few truly ubiquitous technologies and features increasing communication, sensing and processing capabilities. Secondly, human activity recognition plays a key role in the initial visions of ubiquitous computing. Virtually all intelligent interactions envisioned in ubiquitous computing involve inferring what the user is doing at some point. Taking up the challenges offered by activity recognition from mobile phone data could therefore benefit a wide range of researchers.

It is wrong to believe that advances in mobile phone hardware will be sufficient to succeed in the task. Additional inputs do not always have a positive impact on inference and initial experiments with expected sensing technologies do not necessarily generalise well to the mobile phone platform. In order to make full advantage of available context information, new modelling approaches are required. In particular, current inference models have difficulties with missing data, ignore obvious long-range and interpersonal dependencies in human behaviour and are sensitive to temporal variations.

Ultimately, the development of accurate human activity recognition using mobile phones is conditioned upon the public availability of large and reliable datasets. In that respect, examples of collaborations with mobile phone operators for data collection are still very rare. Also, unlike researchers in certain other fields, researchers in mobile-phone-based activity recognition are not yet accustomed to making their evaluation data available to other researchers for comparison.

REFERENCES

Anderson, I. and Muller, H. 2006a. Context awareness via GSM signal strength fluctuation. In *Pervasive '06: Proceedings of the 4th International Conference on Pervasive Computing, Late breaking results*. Oesterreichische Computer Gesellschaft, 27–31.

Anderson, I. and Muller, H. 2006b. Practical activity recognition using GSM data. Tech. Rep. CSTR-06-016, Department of Computer Science, University of Bristol. July.

Antifakos, S., Michahelles, F., and Schiele, B. 2002. Proactive instructions for furniture assembly. In *UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing*. Springer-Verlag, London, UK, 351–360.

Apple, Inc. 2006. Nike + Apple. `http://www.apple.com/ipod/nike/`.

Apple, Inc. 2009. Apple iphone. `http://www.apple.com/iphone/`.

Bao, L. and Intille, S. S. 2004. Activity recognition from user-annotated acceleration data. In *Pervasive '04: Proceedings of the 2nd International Conference on Pervasive Computing*. Springer-Verlag GmbH, 1–17.

Barfield, W. and Caudell, T. 2001. Boeing's wire bundle experiment. In *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence Erlbaum Associates, 447–467.

Barnard, A. 2009. Growing presence in the courtroom: Cellphone data as witness. `http://www.nytimes.com/2009/07/06/nyregion/06cellphone.html?_r=1`.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics 41,* 1, 164–171.

Bodor, R., Jackson, B., and Papanikolopoulos, N. 2003. Vision-based human tracking and activity recognition. In *Proceedings of the 11th Mediterranean Conference on Control and Automation*. Kostrzewa Joseph, 18–20.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. 1992. A training algorithm for optimal margin classifiers. In *COLT '92: Proceedings of the 5th Annual Workshop on Computational Learning Theory*. ACM, New York, NY, USA, 144–152.

Cho, S.-B., Kim, K.-J., Hwang, K. S., and Song, I.-J. 2007. Anidiary: Daily cartoon-style diary exploits bayesian networks. *Pervasive Computing 6,* 3, 66–75.

Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P., Koscher, K., LaMarca, A., Landay, J. A., LeGrand, L., Lester, J., Rahimi, A., Rea, A., and Wyatt, D. 2008. The mobile sensing platform: An embedded activity recognition system. *Pervasive Computing 7,* 2, 32–41.

Choujaa, D. and Dulay, N. 2009. Activity inference through sequence alignment. In *LoCA '09: Proceedings of the 4th International Symposium on Location and Context Awareness*. Springer-Verlag, Berlin, Heidelberg, 19–36.

Cohen, M. 2009. Text-message marketing. http://www.nytimes.com/2009/09/24/business/smallbusiness/24texting.html.

Dey, A. K. 2001. Understanding and using context. *Personal and Ubiquitous Computing 5,* 1, 4–7.

Dornbush, S., English, J., Oates, T., Segall, Z., and Joshi, A. 2007. XPod: A Human Activity Aware Learning Mobile Music Player. In *Proceedings of the Workshop on Ambient Intelligence, 20th International Joint Conference on Artificial Intelligence (IJCAI-2007)*, 2007 ed.

Eagle, N. and Pentland, A. 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing 10,* 4, 255–268.

Eagle, N. and Pentland, A. 2009. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology* 63, 1057–1066.

Eagle, N., Quinn, J. A., and Clauset, A. 2009. Methodologies for continuous cellular tower data analysis. In *Pervasive '09: Proceedings of the 7th International Conference on Pervasive Computing*. Springer-Verlag, Berlin, Heidelberg, 342–353.

Elliott, G. T. and Tomlinson, B. 2006. Personalsoundtrack: context-aware playlists that adapt to user pace. In *CHI '06 Extended Abstracts on Human factors in Computing Systems*. ACM, New York, NY, USA, 736–741.

enotes.com, Inc. 2009. Activities of daily living evaluation. http://www.enotes.com/nursing-encyclopedia/activities-daily-living-evaluation.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD '96: 2nd International Conference on Knowledge Discovery and Data Mining*, E. Simoudis, J. Han, and U. M. Fayyad, Eds. AAAI Press, 226–231.

Facebook, Inc. 2009. Facebook. http://www.facebook.com/.

Farrahi, K. and Gatica-Perez, D. 2008a. Daily routine classification from mobile phone data. In *Workshop on Machine Learning and Multimodal Interaction (MLMI08)*.

Farrahi, K. and Gatica-Perez, D. 2008b. What did you do today? discovering daily routines from large-scale mobile data. In *ACM International Conference on Multimedia (ACMMM).*

Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting. *Annals of Statistics 28.*

Garris, M., Wilson, C., and Blue, J. 1998. Neural network-based systems for handprint OCR applications. *7,* 8 (August), 1097–1112.

Ghahramani, Z. 1998. Learning dynamic Bayesian networks. In *Adaptive Processing of Sequences and Data Structures*. Springer-Verlag, 168–197.

Gips, J. and Pentland, A. S. 2006. Mapping human networks. In *PERCOM '06: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications*. IEEE Computer Society, Washington, DC, USA, 159–168.

Google, Inc. 2009a. Google. http://www.google.com/.

Google, Inc. 2009b. Google latitude. http://www.google.com/intl/en_us/latitude/intro.html.

Google, Inc. 2009c. Google voice. http://www.google.com/googlevoice/about.html.

Groundspeak, I. 2000. Whereigo. http://www.wherigo.com/.

Gupta, M., Intille, S. S., and Larson, K. 2009. Adding GPS-control to traditional thermostats: An exploration of potential energy savings and design challenges. In *Pervasive '09: Proceedings of the 7th International Conference on Pervasive Computing* (2009-05-06). Lecture Notes in Computer Science, vol. 5538. Springer, 95–114.

GYÖRBÍRÓ, N., FÁBIÁN, A., AND HOMÁNYI, G. 2009. An activity recognition system for mobile phones. *Mob. Netw. Appl. 14,* 1, 82–91.

HAN, J. AND BHANU, B. 2005. Human activity recognition in thermal infrared imagery. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern - Workshops*. IEEE Computer Society, Washington, DC, USA, 17.

HARVILLE, M. AND LI, D. 2004. Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera. In *CVPR '04: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision*. Vol. 2. 398–405.

HEWLETT-PACKARD DEVELOPMENT COMPANY, L. 2008. Mscape. `http://www.mscapers.com/`.

HIGHTOWER, J. AND BORRIELLO, G. 2001. Location systems for ubiquitous computing. *Computer 34,* 8, 57–66.

HUANG, W., NAKAMORI, Y., AND WANG, S.-Y. 2005. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research 32,* 10, 2513–2522.

JAIN, A. K., MAO, J., AND MOHIUDDIN, K. M. 1996. Artificial neural networks: a tutorial. *Computer 29,* 3, 31–44.

KATZ, S. 1983. Assessing self-maintenance: Activities of daily living, mobility and instrumental activities of daily living. *Journal of the American Geriatrics Society 31,* 12, 721–726.

KATZ, S., FORD, A., AND ET AL., R. M. 1963. Studies of illness in the aged. the index of ADL: A standardized measure of biological and psychological function. *JAMA: The Journal Of the American Medical Association 185*, 914–919.

KAUTZ, H. 1987. A formal theory of plan recognition. Ph.D. thesis, Dept. of Computer Science, University of Rochester.

KIMMEL, J. 2008. *How do we spend our time?: evidence from the American time use survey*. W. E. Upjohn Institute.

LAMARCA, A., CHAWATHE, Y., CONSOLVO, S., HIGHTOWER, J., SMITH, I. E., SCOTT, J., SOHN, T., HOWARD, J., HUGHES, J., POTTER, F., TABERT, J., POWLEDGE, P., BORRIELLO, G., AND SCHILIT, B. N. 2005. Place Lab: Device positioning using radio beacons in the wild. In *Pervasive '05: Proceedings of the 3rd International Conference on Pervasive Computing*.

LEECH, G., GARSIDE, R., AND BRYANT, M. 1994. CLAWS4: the tagging of the british national corpus. In *Proceedings of the 15th Conference on Computational Linguistics*. Association for Computational Linguistics, Morristown, NJ, USA, 622–628.

LESH, N. AND ETZIONI, O. 1995. A sound and fast goal recognizer. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, 1704–1710.

LESTER, J., CHOUDHURY, T., AND BORRIELLO, G. 2006. A practical approach to recognizing physical activities. In *Pervasive '06: Proceedings of the 4th International Conference on Pervasive Computing*. 1–16.

LIAO, L., PATTERSON, D. J., FOX, D., AND KAUTZ, H. 2007. Learning and inferring transportation routines. *Artificial Intelligence 171,* 5-6, 311–331.

LLOYD, S. 1982. Least squares quantization in PCM. *Information Theory, IEEE Transactions on 28,* 2, 129–137.

MAURER, U., ROWER, A., SMAILAGIC, A., AND SIEWIOREK, D. 2006. Location and activity recognition using eWatch: A wearable sensor platform. In *Ambient Intelligence in Every Day Life*. Springer Press, 86–100.

MAWSTON, N. 2008. Enabling technologies: CMOS beats CCD in half-billion global camera phone market. Report, Strategy Analytics.

MCCOWAN, I., GATICA-PEREZ, D., BENGIO, S., LATHOUD, G., BARNARD, M., AND ZHANG, D. 2005. Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27,* 3, 305–317.

MERRIAM-WEBSTER, INC. 2009. Merriam-Webster online. `http://www.merriam-webster.com/`.

MILUZZO, E., LANE, N., EISENMAN, S., AND CAMPBELL, A. 2007. CenceMe – injecting sensing presence into social networking applications. 1–28.

MILUZZO, E., LANE, N. D., FODOR, K., PETERSON, R., LU, H., MUSOLESI, M., EISENMAN, S. B., ZHENG, X., AND CAMPBELL, A. T. 2008. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *SenSys '08: Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, New York, NY, USA, 337–350.

MySpace, Inc. 2009. MySpace. http://www.myspace.com/.

Nait-Charif, H. and McKenna, S. J. 2004. Activity summarisation and fall detection in a supportive home environment. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on*. Vol. 4. IEEE Computer Society, Washington, DC, USA, 323–326.

Niu, W., Long, J., Han, D., and Wang, Y.-F. 2004. Human activity detection and recognition for video surveillance. In *ICME*. IEEE, 719–722.

Oliver, N. and Horvitz, E. 2004. S-SEER: Selective perception in a multimodal office activity recognition system. In *MLMI '04: Proceedings of the 1st International Workshop on Machine Learning for Multimodal Interaction*. 122–135.

Osmani, V., Balasubramaniam, S., and Botvich, D. 2008. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications 31,* 4, 628–655.

Osuna, E., Freund, R., and Girosi, F. 1997. Training support vector machines: an application to face detection. In *CVPR '97: Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 130–136.

Papliatseyeu, A. and Ibarra, O. M. 2008. Mobile habits: Inferring and predicting user activities with a location-aware smartphone. In *UCAMI '08: Proceedings of the 3rd Symposium of Ubiquitous Computing and Ambient Intelligence*.

Park, H.-S., Yoo, J.-O., and Cho, S.-B. 2006. A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In *FSKD '06: Proceedings of the 3rd International Conference on Fuzzy Systems and Knowledge Discovery*. 970–979.

Park, S. and Kautz, H. 2008. Hierarchical recognition of activities of daily living using multi-scale, multi-perspective vision and RFID. In *Proceedings of the 4th Conference on Intelligent Environments*.

Patterson, D. J., Fox, D., Kautz, H., and Philipose, M. 2005. Fine-grained activity recognition by aggregating abstract object usage. In *ISWC '05: Proceedings of the 9th IEEE International Symposium on Wearable Computers*. IEEE Computer Society, Washington, DC, USA, 44–51.

Patterson, D. J., Liao, L., Gajos, K., Collier, M., Livic, N., Olson, K., Wang, S., Fox, D., and Kautz, H. 2004. Opportunity knocks: A system to provide cognitive assistance with transportation services. In *UbiComp '04: Proceedings of the 6th International Conference on Ubiquitous Computing*. 433–450.

Pearl, J. 1985. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*. 329–334.

Pentland, A. S. 2007. Automatic mapping and modeling of human networks. *Physica A: Statistical Mechanics and its Applications 378,* 1 (May), 59–67.

Perkowitz, M., Philipose, M., Fishkin, K., and Patterson, D. J. 2004. Mining models of human activities from the web. In *WWW '04: Proceedings of the 13th International Conference on World Wide Web*. ACM, New York, NY, USA, 573–582.

Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., and Hahnel, D. 2004. Inferring activities from interactions with objects. *IEEE Pervasive Computing 3,* 4, 50–57.

Preuveneers, D. and Berbers, Y. 2008. Mobile phones assisting with health self-care: a diabetes case study. In *MobileHCI '08: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 177–186.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*, 1 ed. Morgan Kaufmann.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*. 257–286.

Raento, M., Oulasvirta, A., Petit, R., and Toivonen, H. 2005. Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing 4,* 2, 51–59.

Rao, L. 2009. Twitter flew above the 50 million uniques mark for the first time in july. http://www.techcrunch.com/2009/08/25/

twitter-flew-above-the-50-million-uniques-mark-for-the-first-time-in-july/.

Reddy, S., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. 2008. Determining transportation mode on mobile phones. In *ISWC '08: Proceedings of the 12th IEEE International Symposium on Wearable Computers*. 25–28.

Robertson, N. and Reid, I. 2006. A general method for human activity recognition in video. *Computer Vision and Image Understanding 104,* 2, 232–248.

Rowley, H. A., Baluja, S., and Kanade, T. 1998. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20,* 1, 23–38.

Schmidt, C. F., Sridharan, N. S., and Goodson, J. L. 1978. The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence 11,* 1-2, 45–83.

Shachtman, N. 2003. A spy machine of DARPA's dreams. http://www.wired.com/techbiz/media/news/2003/05/58909?currentPage=2.

Sohn, T., Varshavsky, A., Lamarca, A., Chen, M., Choudhury, T., Smith, I., Consolvo, S., Hightower, J., Griswold, W., and de Lara, E. 2006. Mobility detection using everyday GSM traces. 212–224.

Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., and Tröster, G. 2008. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing 7,* 2, 42–50.

Tapia, E. M., Intille, S. S., and Larson, K. 2004. Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive '04: Proceedings of the 2nd International Conference on Pervasive Computing*. 158–175.

The Associated Press. 2006. Companies discover marketing power of text messaging. http://www.cbc.ca/consumer/story/2006/09/25/text-messaging.html.

Tryhorn, C. 2009. Nice talking to you ... mobile phone use passes milestone. http://www.guardian.co.uk/technology/2009/mar/03/mobile-phones1.

Turk, M. A. and Pentland, A. P. 1991. Face recognition using eigenfaces. In *CVPR '91: Proceedings of the 1991 IEEE Conference on Computer Vision and Pattern Recognition*. 586–591.

Twitter, Inc. 2006. Twitter. http://twitter.com/.

U.S. Bureau of Labor Statistics. 2009. American time use survey. http://www.bls.gov/tus/.

Vemuri, S. and Bender, W. 2004. Next-generation personal memory aids. *BT Technology Journal 22,* 4, 125–138.

Viola, P. and Jones, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision 57,* 2, 137–154.

Viterbi, A. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory 13,* 2, 260–269.

Wallhoff, F., Zobl, M., and Rigoll, G. 2004. Action segmentation and recognition in meeting room scenarios. In *ICIP '04: Proceedings of the 2004 International Conference on Image Processing*. 2223–2226.

Ward, J. A., Lukowicz, P., Tröster, G., and Starner, T. 2006. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 1553–1567.

Weiser, M. 1991. The computer for the 21st century. *Scientific American*.

Williams, A., Ganesan, D., and Hanson, A. 2007. Aging in place: fall detection and localization in a distributed smart camera network. In *ACM-MM '07: Proceedings of the 15th International Conference on Multimedia*. ACM, New York, NY, USA, 892–901.

Witten, I. H. and Frank, E. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann.

Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., and Rehg, J. M. 2007. A scalable approach to activity recognition based on object use. In ICCV '07: Proceedings of the 11th International Conference on Computer Vision. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 1–8.

Yeo, J., Kotz, D., and Henderson, T. 2006. CRAWDAD: a community resource for archiving wireless data at dartmouth. *SIGCOMM Computer Commununication Review 36,* 2, 21–22.

Zhang, K., Li, H., Torkkola, K., and Gardner, M. 2007. Adaptive learning of semantic locations and routes. In *LoCA '07: Proceedings of the 3rd International Conference on Location- and Context-Awareness*. 193–210.

...