2010 International Conference on Body Sensor Networks

Multisensor Fusion in Smartphones for Lifestyle Monitoring

Raghu K. Ganti
Department of Computer Science
University of Illinois
Urbana, IL, 61801
Email: rganti2@cs.uiuc.edu

Soundararajan Srinivasan, and Aca Gacic
Robert Bosch LLC
Pittsburgh, PA 15212
E-mail: Soundar.Srinivasan@us.bosch.com,
Aca.Gacic@us.bosch.com

***Abstract*—Smartphones with diverse sensing capabilities are becoming widely available and pervasive in use. With the phone becoming a mobile personal computer, integrated applications can use multi-sensory data to derive information about the user's actions and the context in which these actions occur. This paper develops a novel method to assess daily living patterns using a smartphone equipped with microphones and inertial sensors. We develop a feature-space combination approach for fusion of information from sensors sampled at different rates and present a computationally light-weight algorithm to identify various high level activities. Preliminary results from an initial deployment among eight users indicate the potential for accurate, context-aware, and personalized sensing.**

***Index Terms*—Mobile computing, Algorithm design and analysis, Ubiquitous computing, Wearable computers, Activity identification, Activities of daily living.**

I. Introduction

Smartphones, with a wide array of sensing devices, are increasingly becoming common these days. It has been noted in [18] that an individual's cellphone is a personal proxy, a context aware device, an activity inference device, and a payment proxy. A mobile phone is increasingly becoming popular as a personal computer.

The main challenge in detection of context and activity using a smartphone comes from the fact that it is not primarily designed to collect and infer activity related information. For that purpose, custom-made mobile sensor platforms are built and used for identification of context related to human activity and are optimized to boost classification performance [13], [17], [16]. In practice, however, the main challenge is the usability of the chosen solution. Unless the user is already familiar with the technology and/or it provides a distinct, unique value, the adherence almost always declines rapidly after the initial period and the user eventually discontinues to use it.

Because of the proven value to the user, smartphones present a natural platform for hosting context sensitive applications. Analogous to a user owning a personal computer who is more likely to install new software rather than buy specialized devices for each application, a common user who already owns a cellphone is more likely to adopt a software for a new application rather than buy a specialized device for achieving this purpose. Nevertheless, the range of available embedded sensors and the practical use of these devices in the real life pose serious challenges to the developer of robust context detection algorithms.

With Software Development Kits (SDKs) such as Symbian S60 [11] and the iPhone SDK available currently, third party applications and software for the smartphones can be more easily developed and made available to the common user. The vision of the cellphone being a proxy for the user and their activities will soon become a reality.

In addition to microphones and digital cameras, today's smartphones are often equipped with micro-electromechanical (MEMS) sensors, mainly accelerometers, which have a small form factor and low power consumption advantages without compromising on performance. As technology matures and more sensors get integrated in smartphones, we anticipate that an emerging category of applications will exploit the abundance of available data and apply multi-sensor data fusion to deliver reliable services that depend on understanding the context in which they are used.

In this article, we take a step in the direction of designing these robust applications in the face of the aforementioned challenges. We specifically address the problem of the cellphone as an activity inference device. Human activity identification will provide input to a rich set of new applications in the realm of healthcare, personal record keeping, entertainment, and safety. The rest of the paper is organized into five sections. Section II presents the implementation details of our system (data collection and storage). Section III presents our algorithm for inferring human activities and Section IV summarizes the results from our deployment. We discuss related work in Section V and conclude with discussion of various application scenarios and directions for future work in Section VI.

II. System Design

Smartphones are quite prevalent these days and their capabilities have also increased multifold in the past few years. Examples of such smartphones equipped with various sensors include the Nokia N-series (N82, N95, N96), Apple iPhone, and the BlackBerry. Many of these smartphones are equipped with location, motion, light, audio, and video sensors. Since Nokia provides a large number of smartphones that have a

This work was done when Raghu Ganti was an intern at Robert Bosch LLC

978-0-7695-4065-8/10 $26.00 © 2010 IEEE
DOI 10.1109/BSN.2010.10

IEEE computer society

common operating system and APIs, we choose to use the Nokia N95 for our work.

We design and implement a general software architecture for the purpose of data collection on the Nokia N95. The N series of Nokia phones use a client-server based operating system, the Symbian OS [11], designed for resource constrained mobile devices. Access to lower level hardware is provided through a request callback sequence, where servers (abstractions of lower level hardware) respond to requests from clients.

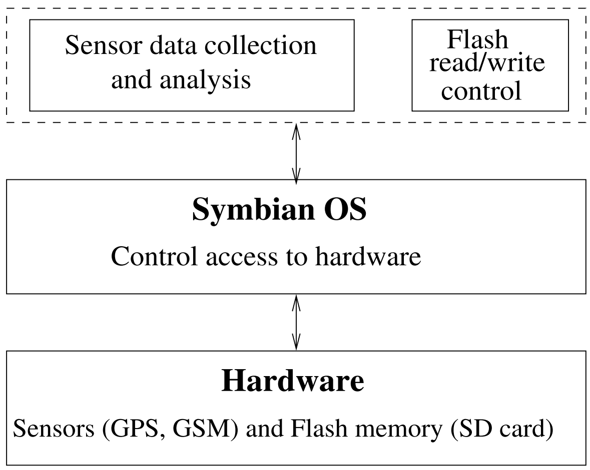


Fig. 1. Figure depicting the various components of our software design on the cellphone using Symbian OS.

Fig. 2. Figure showing the application for activity tagging

Figure 1 shows the components of our software architecture, which enables a generic and flexible collection of data from various sensors. We can see from Figure 1 that the components fall into three categories, the lowest level includes the hardware of the cellphone (microphone, GPS, accelerometer), abstractions of which are provided by the server components of Symbian OS. The modules in the application layer (top most) provide three main functionalities: (i) Data collection from the server components, (ii) Recording data collected from the sensors, and (iii) Reading data recorded from the flash for upload to a PC for data analysis. Our architecture is modular, flexible, and extensible and enables data collection from various sensors with ease. A snapshot of the data collection application that utilizes the above architecture is shown in Figure 2. The application allows for tagging the data streams being recorded with the corresponding activity (chosen by the user from the drop down list).

Our current implementation records four different sensors: the microphone, accelerometer, GPS, and GSM (GSM information is used to determine the user's location when GPS signals are unavailable) for offline analysis. While our current inference algorithm is fairly light-weight, the initial development focuses on a proof-of-concept implementation of multisensor fusion for activity detection. We are currently working on enabling such a real-time activity inference technique on the phone. Further, recording the identified activity may be useful for long-term trend analysis, as is shown by *smart attire* [5].

III. Activity Inference

In this section, we develop the method for identifying activities termed as Activities of Daily Living (ADL). ADLs [10], [14] are of importance, especially in the medical community, where they are used to ascertain the health of an elderly individual. In this paper, we show that integrating sensor data from the microphone and the acceleration sensor embedded in the N95 is a promising approach for ADL monitoring. We develop a feature-space-combination approach, in which we extract information from both sensors sampled at different rates. This is accomplished by a synchronous feature extraction approach in which features from each sensor are computed independently at the same, constant time-frame rate. The extracted features are fed to a computationally light-weight algorithm, suitable for implementation on a smartphone, such as the N95.

In our study, the user wears the phone on the waist and performs an activity (e.g. cooking). Before the start of the activity, the user activates the data capture application within the smartphone (as shown in Section II). The data collection modules samples the acceleration sensor at 7 Hz and the microphone at 8 kHz. The low sampling rate of the acceleration sensor is due to the Symbian OS (and we have no control over it). The above sensor signals form the input to an ADL monitor. In the current solution, data processing is done offline to facilitate rapid development of the algorithm.

We conducted an empirical study to obtain the training and testing data set for the automated classifier. Eight distinct ADLs and instrumental activities of daily living (IADL), as shown in Table I, are considered in the present study. These eight activities are chosen as they form the basic ADLs and IADLS [10], [14]. Note that the activities in Table I comprise of both static and dynamic activities. An activity is "static"when there is no significant acceleration signal detected at the waist while performing that activity. A "dynamic" activity, on the other hand, contributes to significant acceleration while accomplishing the activity. Hence, we can see that not all of the activities in Table I can be differentiated based on acceleration signals alone.

We use an hidden Markov Model (HMM), to model each activity. Since HMMs are well known models of time series and have been used successfully to recognize human actions [5], we use it in the present study to identify ADLs/IADLs. We extract the following features from the acceleration data in overlapping time frames of 5 seconds with a 1.67 seconds frame shift. Let us assume that the acceleration signal is $a(t)$, which is discrete. The identified time window is $[t_1, t_2]$, for which features are extracted from the signal. Further, the acceleration signal has three components at any given time (t), which are the acceleration in x-axis (a_x), y-axis (a_y), and z-axis (a_z). We compute the features across each axis. The first feature is the relative change in body orientation (O) in xy-z plane with respect to a calibration phase when the user is presumed to be standing. This is defined as follows:

$$\theta = \arctan(\frac{\sqrt{((m_{a_x})^2 + (m_{a_y})^2)}}{m_{a_z}} \quad (1)$$

In Equation 1, m_{a_x} is the mean of x axis acceleration, m_{a_y} is the mean of the y axis acceleration, and m_{a_z} is the mean of the z axis acceleration.

The next feature is related to the energy expended in a particular physical activity, which is measured as the energy of the acceleration signal in the given time window for all the three axes. It is specified as follows:

$$EE = mean(\sum_{x,y,z}(a^2)) \quad (2)$$

We also compute the skewness of the magnitude of the 3-dimensional acceleration, given by the following equation:

$$SK = E[(a_i - \mu)^3/\sigma^3] \quad (3)$$

In Equation 3, $\mu = mean(a_i)$, σ is the standard deviation of a, and $E[t]$ represents the expected value of the quantity of t. Finally, we compute the entropy of the acceleration, which is given by:

$$H(a) = -\sum_{i=t_1}^{t_2} p(a_i) log_2 p(a_i) \quad (4)$$

In Equation 4, $p(.)$ is the probability mass function of a.

Relative inclination helps in distinguishing activities that depend on whether a person is sitting (e.g., "Eating"), standing (e.g., "Cooking"), and lying-down (e.g., "Watching TV"). Energy expenditure, skewness, and entropy help distinguish between dynamic activities (e.g., "Aerobic) and static ones (e.g., "Desk Work"). Figure III shows an example of how the vector magnitude is different between "Aerobic" and "Desk Work."

The microphone data is also processed at the same frame length. Let us assume that this is represented by $s(t)$ and the time duration for which the signal is processed is $[t_1, t_2]$. We extract spectral shape features that are related to the audio content [7]. The *cepstral coefficients (c)* are defined as follows:

$$\begin{aligned} f &= fft(s, n) \\ x &= melbank(p, n, fs) \\ n_2 &= 1 + \lfloor n/2 \rfloor \\ z &= log(x * |f(1 : n_2)^2|) \\ c &= DCT(z) \end{aligned}$$

In the above equations, fft performs the Fourier transform (FFT), n is the number of points for the FFT, $melbankm$ computes the mel-spaced filterbank and DCT computes the discrete cosine transform of z. We set $p = 26$, $n = 256$, and $fs = 8000$. These "cepstral coefficients" help in differentiating between different classes of dynamic activities (e.g., "Cooking" and "Hygiene"), or different classes of static activities (e.g., "Meeting" and "Driving"). We use the first 12 cepstral coefficients as they are well known to be the most useful in describing the content of an audio signal [7]. Figure III shows an example of how the spectral shape, computed using cepstral coefficients, is significantly different between "Cooking" and "Driving".

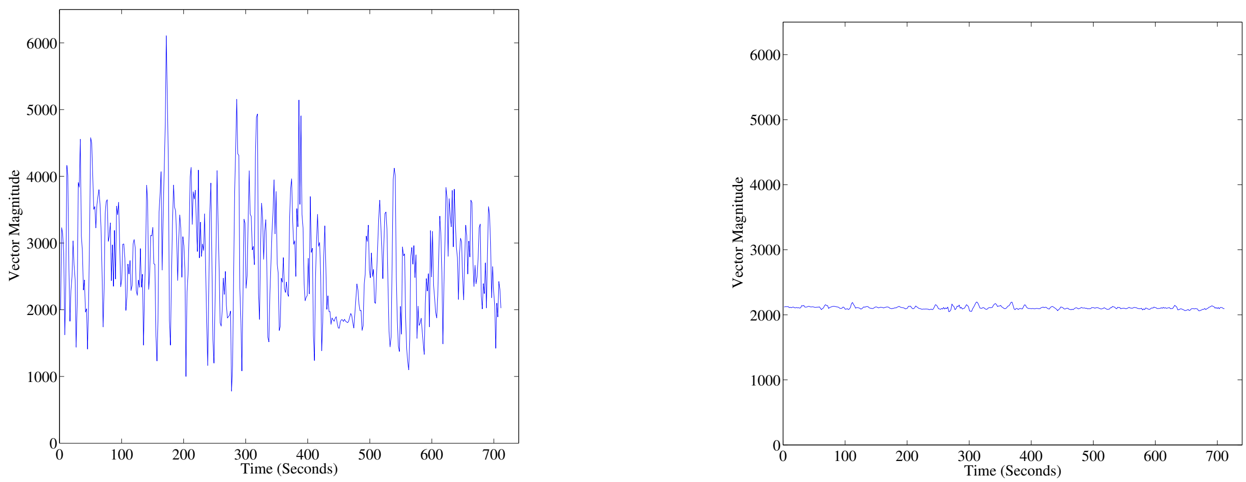
IV. RESULTS

We now present the descriptive results of our empirical study (see Section III) in this section. We recruited eight male participants between the age groups of 20-37 years to participate as subjects for the empirical data collection experiments. We instructed the users to go about their regular routine and perform their daily activities as naturally as possible and placed no restrictions on the location or time of the day. Users were encouraged to wear the device as much as possible for a period of eight weeks in either their pocket or a carrying pouch. We compensated each participant with a $20 gift card to a local movie theater. All participants signed an user agreement that stated that they agree to the collection and use of microphone, acceleration, GPS (if available), and GSM (cell information that can be used to track location) data for research purposes.

In order to collect data to train the classification algorithm and validate the results of classification, we modified the data collection part of our software design (Section II) to enable users to label their activities. Specifically, we instructed the users to label the beginning and the end of each activity. Additionally, when the phone is switched on, the user calibrated the

Activity		
Name	Type	Activity Description Reference
Aerobic	Dynamic	Walking, running, lifting weights, etc.
Cooking	Dynamic	Food preparation, heating, grilling, etc.
Desk Work	Static	Typing, reading at desk.
Driving	Static	Driving in a car.
Eating	Static	Eating while seated or standing.
Hygiene	Dynamic	Washing dishes, brushing teeth, etc.
Meeting	Static	Present in, or attend a meeting.
Watching TV	Static	Watching TV while not performing any of the above activities.

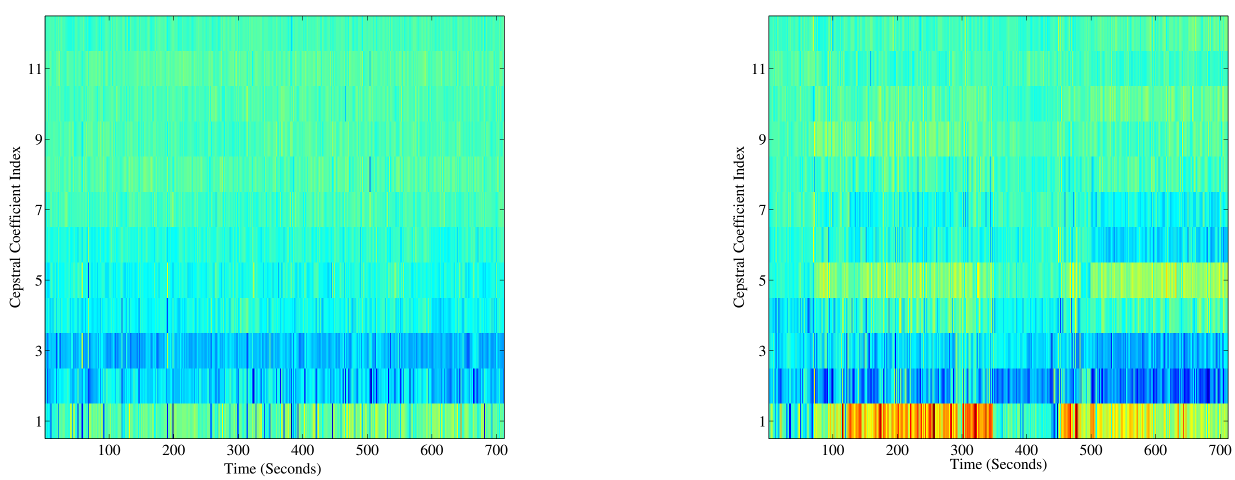
TABLE I
ACTIVITIES OF DAILY LIVING CONSIDERED WITHIN THE EMPIRICAL STUDY



(a) Vector magnitude of triaxial acceleration during "Aerobic" activity.

(b) Vector magnitude of triaxial acceleration during "Desk Work."

Fig. 3. Figure illustrating how the vector magnitude feature helps differentiate between dynamic and static activities.



(a) Cepstral coefficients of an instance of "Cooking."

(b) Cepstral coefficients of an instance of "Driving."

Fig. 4. Figure illustrating how the cepstral coefficients help differentiate between different static activities.

acceleration axes by standing still for a period of 10 seconds. The start and end of calibration was cued by making the device vibrate.

A total of 80 hours of tagged activity data was collected. A partial data set of 45 hours is used for developing and training the automated classification algorithm. Eight activity-level HMMs are trained, one for each activity listed in Table I. All have 3 states, whose output distribution is modeled as a mixture of 8 Gaussians. The 3 state model is chosen to model the 'transition-into', the 'steady state', and the 'transition-out' of each activity. Testing is performed on a 7 hour subset of the remaining data (i.e., data not including in training). The rest 28 hours of data is unusable because the users did not either calibrate the device or label the activities correctly. An HMM toolkit, HTK [20], is used for training. During testing, we perform a maximum-likelihood decoding to determine the most likely activity. This form of decoding could be viewed as a single finite state model composed of individual HMMs with transitions between various activities classes modeled as equally likely. We make this simplifying assumption currently due to lack of a large dataset to model transition between activities accurately. Note that recognition is user independent; the data from all the users are used to construct the HMMs and the testing does not exploit the knowledge of the user identity.

The results, in terms of accuracy of classification, are summarized in Table II. Accuracy refers to the portion of time windows in which the classified and labeled activities match. The results indicate that the recognition accuracy is quite high for the following activities: "Aerobic,""Cooking," "Driving," and "Hygiene'.' We find that "Eating" is hard to distinguish from "Watching TV" because a majority of "Eating" activity was performed when the user was "Watching TV" and hence the audio and acceleration features are quite similar. One way to alleviate this would be allow for "N-best" outputs from the classifier, where "N" refers to those outputs during decoding that exceed a threshold on their likelihoods. The low accuracy of "Meeting" on the other hand is due to the availability of a limited amount of our data corresponding to this class. Table III shows the results in terms of precision and recall. We observe from Table III that our precision results are also quite good, except for the "Meeting" class due to the limitation mentioned above. Overall, the results in Table II and Table III show the potential of combining the information from the acceleration sensor and the microphone for the identification of ADLs. However, further evaluation is required to confirm the statistical validity and significance of these preliminary results.

We now present the confusion matrix in Table IV. As we mentioned earlier, our dataset consists of people "Eating" while "Watching TV", and hence both these activities were confused with each other. We also observe that "Desk Work" and "Meeting" are confused with each other due to the open space work environment in which these data were collected.

Finally, we provide an empirical justification regarding our choice of the 3-state HMM. Table II also shows the accuracy results when using a Gaussian Mixutre Model (a 1-state HMM) or a 5-state HMM for comparison. We can observe from Table II that in terms of the average accuracy, the 3-state HMM outperforms the 1-state and 5-state HMMs, thus supporting our choice of using the "transition-into", "steady", and "transition-out" states. We conclude that our choice of discriminative acceleration and audio features and a 3-state HMM provide a promising approach to identifying ADLs.

Activity	**Accuracy (%) 3-state**	**Accuracy (%) 1-state**	**Accuracy (%) 5-state**
Aerobic	82	79.3	83.1
Cooking	100	100	100
Desk Work	53	34	50
Driving	87.6	96	77
Eating	12.7	12.7	14
Hygiene	99	64	43
Meeting	12.7	12.7	14
Watching TV	88	87	88

TABLE II
PERFORMANCE OF THE AUTOMATED ADL CLASSIFIER

Activity	**Precision %**	**Recall %**
Aerobic	76.8	81.9
Cooking	76.4	100
Desk Work	50.8	53
Driving	100	87.6
Eating	51.2	12.7
Hygiene	65.6	99
Meeting	12.5	12.7
Watching TV	55.9	88

TABLE III
PRECISION AND RECALL OF THE CLASSIFIER

V. RELATED WORK

A Gaussian Mixture Model (GMM) based approach combined with a finite state machine is developed in [6] for the purpose of identification of early morning bathroom activities, such as washing face, brushing teeth, and shaving. An accelerometer strapped to the wrist is used for activity identification. We identify a broader set of activities and use a cellphone for identifying these activities.

RFIDs are used to identify ADLs in [13]. The authors build a system called Proact that uses inputs from RFIDs attached to various objects and a reader attached to a glove. These inputs are used to create models using *dynamic Bayesian networks*, that are then used for the identification of the activities. In [17], an approach that augments the use of RFIDs with an accelerometer mounted on the glove with the RFID reader is presented. In this work, RFID tag readings are used to narrow down the set of activities based on the type of object being used. In contrast to the above paper, our work uses existing sensors from cellphones and avoids the use of cumbersome devices like a glove with an RFID reader attached and RFID tags.

Logan and Healey present an approach for identifying a minimal set of sensors for recognizing eating and meal preparation [8]. An adaptive boosting (AdaBoost) classifier

Activity	Aerobic	Cook	Desk	Drive	Eat	Hygiene	Meet	TV
Aerobic	1197	145	0	0	120	0	0	0
Cooking	0	523	0	0	0	0	0	0
Desk Work	329	0	958	0	0	0	519	0
Driving	33	0	0	891	7	0	0	86
Eating	0	0	333	0	241	199	0	1117
Hygiene	0	0	0	0	4	380	0	0
Meeting	0	0	596	0	0	0	87	0
Watching TV	0	17	0	0	99	0	91	1523

TABLE IV
TABLE SHOWING THE CONFUSION MATRIX

is used for separating eating and associated tasks from other activities. Our work addresses the identification of a larger set of activities with a limited set of sensors that are available on the cellphone.

Zhen et al., use state change sensors installed on various household objects such as doors, drawers, and refrigerators for the purpose of the identification of activities such as cooking, shopping, washing, and bathing [21]. A self-adaptive neural network called "Growing Self-Organizing Maps" is used for activity identification. In contrast to our work, the above paper used inputs from objects tagged with state change sensors.

A specialized device that records various sensor readings, such as the microphone, light, accelerometer, and barometer is developed in [1]. An on-device inference algorithm that identifies activities such as walking, sitting, climbing stairs, and brushing teeth is also presented. Our work, on the other hand, identifies a broader set of activities and does not require a specialized device to be used. Further, the set of features computed in the paper is quite large.

Cell phone based basic activity identification has been done earlier [9], [15], [3]. Complementary to such applications, our work addresses the question of identifying complex activities: the ADLs.

VI. DISCUSSION AND CONCLUSION

We have developed a Bayesian learning based approach for identifying ADLs using commonly available smartphones. Our system uses two sensor modalities, the microphone and the accelerometer to achieve activity inference. We show through extensive experimental studies conducted across eight people over eight weeks that our approach to activity inference is quite robust and accurate. Note that our approach does not require each new user to train the system. However, the performance can be improved by adapting the models to individual users. This can be accomplished by an unsupervised adaptation of the HMM models to the new user data [19]. Our work can also be viewed as context recognition using commonly available sensors in today's smartphones without introducing any additional equipment. Assuming that the location of the user can be determined via the recorded GPS and GSM data, we focused on inferring the user's activity as an additional indicator of context [12]. I

Although the results are promising, a few major practical challenges came up during data collection and testing of the system. First, we have to collect more data to address the following limitations: the data was collected in sparse bursts and unevenly distributed between the various classes. Second, there are many situations in which the phone is not normally carried by the user, most notably at home (e.g., in the bathroom) where many ADLs are performed. Third, the recorded microphone data varied significantly between the users who were wearing the phone in the pocket versus the users who used the carrying pouch. Finally, the sampling of the acceleration data was done asynchronously and we noticed that the sampling rate was affected by the sampling rate of the microphone signal. This is a peculiarity of the Symbian OS on the Nokia N95 phone since acceleration sampling is implemented as a low priority task.

Our future work will address a more general concept of sensor data fusion to improve context detection. In the first step, we will use GPS and GSM data to determine the locality of the user. Coarse-grained localization using GSM will be used to boost the activity recognition performance and provide logical maps of user's location whereas GPS data will be used to deliver more precise physical locality information for applications that require such information, such as route planning and tracking for fitness applications.

We will explore several application domains where such a system could find use:

1) Medical monitoring to support behavior management
2) Management of mode of operation of smartphones and user profiles on digital mobile devices based on detected context.
3) Context-rich activity monitoring and feedback to promote behavior change and enable new ways of social networking.

Smartphones are already being used in applications related to health management, mostly either as data aggregation and gateway devices or as health related content delivery devices, or both. One powerful advantage of smartphones is that they can deliver the content at the right place and time, which makes them suitable for behavior modification applications. This is especially interesting in chronic care management where environment and psychosocial conditions play a major role in successful treatment [4]. Compliance is the main target of behavior modification techniques. Medical monitoring applications can, for example, provide more effective reminder solutions that take into account the physical

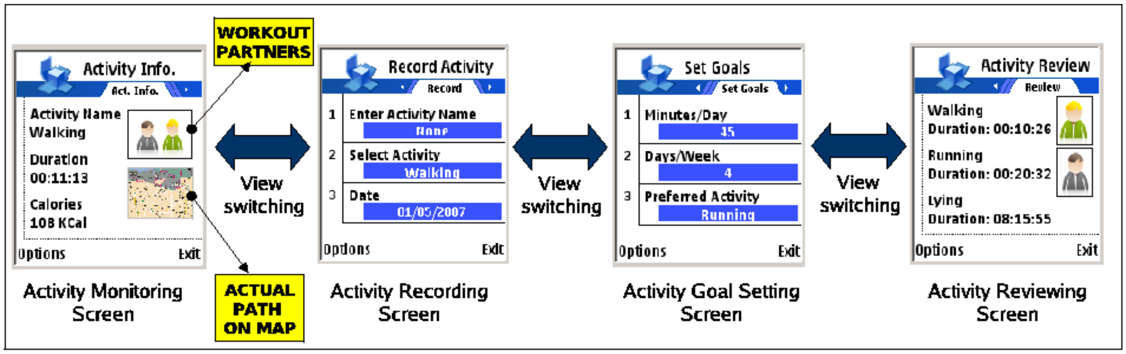


Fig. 5. Snapshot of the four categorical screens of an example fitness application

and cognitive state and the social environment of the user. Smartphones alone will not gather all necessary information needed to understand behavior models. However, they will serve as a powerful component of future behavior management solutions. On the one hand, smartphones have the capability of discerning between many physical and social activities, small set of which we demonstrated in the work done so far. On the other hand, they will also be used to deliver medical content and prompt users for actions to close the monitoring-feedback loop. In addition, we will explore applications that log events leading to and during incidents (e.g., falls) resulting in physical injury to understand better the circumstances under which these incidents happen.

Context detection allows managing notifications and reminders and other services on smartphones in a better way by taking into consideration the environment and the activity of the user within that environment. Knowing the context in which the phone is used allows us to determine, for example, whether or not the user can be interrupted, or if the ring volume should be turned up in a noisy environment or turned off in a meeting. Context can be used to automatically discover and consume available services that are beneficial at a certain time and place. For example, if the phone discovers that the user is driving, it may search for a hands-free car service to automatically subscribe to under a defined condition, such as the number of rings, or triggered by the detection of the phone being fumbled or dropped. Location context using GPS and GSM information opens additional possibilities such as switching to the silent mode in public buildings or alerting a jogger if she veers off a planned route.

Activity and location tracking can be used to provide a support for a variety of fitness and wellness behavior change applications that will make use of real-time feedback that smartphones can provide. In addition, smartphones are suitable platforms for social networking applications because of their ubiquitous presence and nomadic use. These two aspects offer a unique way of promoting behavior change through social influence. For example, one study has shown that social networks play a significant role in spreading obesity [2]. We envision applications that share activity information between users in a social network to promote positive reinforcement within a group. These groups could either be already existing, such as family and close friends, or can be organically formed by enabling users to discover shared interests and goals. Smartphones will have the capability of scanning the social neighborhood simultaneously with recording personal activities. That way the users will be able to interact with the groups of users who perform activities at same time and place, which often results in reinforcing influence.

Figure 5 shows an example of one such application domain related to personal fitness. In order to provide comprehensive engagement and feedback information, we propose that four specific operations be available to the user. They are monitoring current activity information, recording past activity information, recording future goals and reviewing past history of activities. A context-sensitive menu at the bottom of the screen provides different menu options based on the current view that is active. The “Activity Info.” screen allows users to view a list of features including activity name, date, begin time, end time, duration, calories (energy expenditure), total distance, average speed and step count. The “Record Activity” screen allows users to record or tag additional information about past activity by either selecting from a list of existing activities or by entering a new activity name. The “Set Goals” screen allows users to set personal goals for the future such as the number of minutes that they would like to work out each day and the frequency in terms of days per week. The user can also set personal barriers that prohibit them from performing physical activity and personal goals that they would like to accomplish by performing the activity, which could be utilized to provide motivational feedback. The “Activity Review” screen allows users to review historical activity information either in chronological order or by searching using specific query criteria. While the “Activity Info.” screen provides real-time feedback, the “Activity Review” screen is used to provide reflective feedback to the user. This helps the user to track progress toward personal goals. The application can provide context-related information, such as workout partners and the actual path covered on a map while performing an activity, using the “Activity Info.” and “Activity Review” screens on the smartphone's user interface.

We will not only investigate technical solutions but also address practical challenges in these applications by investigating usability and human factors aspects. We plan to conduct user studies and develop tools for assessment of the efficacy of such systems. Future mobile computing devices continue to be an essential tool for communication, entertainment, and delivery of various services which can considerably benefit from the ability to adapt to location, nearby social network, resources available in the proximity of the user, and the user's activity. The range of available embedded sensors is likely to expand and allow development of these applications with higher reliability.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments that helped improve this paper. Raghu K. Ganti's work was funded in part by NSF grants CNS 09-05014 and CNS 06-26342.

REFERENCES

[1] T. Choudhury et al. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2):32–41, April-June 2008.

[2] N. A. A. Christakis and J. H. H. Fowler. The spread of obesity in a large social network over 32 years. *N Engl J Med*, 357(4):370–379, July 2007.

[3] T. Denning et al. Balance: Towards a usable pervasive wellness application with accurate activity inference. In *HotMobile '09*, pages 1–6, 2009.

[4] E. B. Gallagher and T. D. Straton. Health behavior in persons living with chronic conditions. In D. S. Gochman, editor, *Handbook of Health Behavior Research I (1st edition)*. Springer, Upper Saddle River, NJ, USA, 1997.

[5] R. K. Ganti, P. Jayachandran, T. F. Abdelzaher, and J. A. Stankovic. Satire: a software architecture for smart attire. In *Proc. of ACM MobiSys*, pages 110–123, 2006.

[6] N. F. Ince, C.-H. Min, and A. H. Tewfik. A feature combination approach for the detection of early bathroom activities with wireless sensors. In *Proceedings of HealthNet'07*, pages 61–63, 2007.

[7] D. Li, I. Sethi, N. Dimitrova, and T. McGee. Classification of general audio data for content-based retrieval. *Pattern Recognition Letters*, 22(5):533–544, 2001.

[8] B. Logan and J. Healey. Sensors to detect the activities of daily living. In *Proceedings of EMBS Annual Conference*, pages 5362–5365, 2006.

[9] E. Miluzzo et al. Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application. In *Proc. of SenSys*, November 2008.

[10] New York State Office of the Aging - Toolkit for Caregivers. http://www.aging.ny.gov/Caregiving/Toolkit/7CaringforYourParentsFactSheetsinEnglish/ChecklistofActivitiesofDailyLiving.pdf.

[11] Nokia S60 SDK. http://www.forum.nokia.com/.

[12] A. Ofstad, E. Nicholas, R. Szcodronski, and R. R. Choudhury. Aampl: accelerometer augmented mobile phone localization. In *MELT*, pages 13–18, 2008.

[13] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):10–17, October-December 2004.

[14] Polisher Research Institute . http://www.abramsoncenter.org/PRI/documents/IADL.pdf.

[15] S. Saponas et al. ilearn on the iphone: Real-time human activity classification on commodity mobile phones. Cse tech report, DUB, University of Washington, 2008.

[16] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger, J. Shaffer, and F. L. Wong. Sensay: a context-aware mobile phone. In *Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on*, pages 248–249, 2003.

[17] M. Stikic, T. Huynh, K. V. Laerhoven, and B. Schiele. Adl recognition based on the combination of rfid and accelerometer sensing. In *Proceedings of Pervasive Health*, pages 258–263, 2008.

[18] R. Want. You are your cell phone. *IEEE Pervasive Computing*, 7(2):2–4, April–June 2008.

[19] P. Woodland, D. Pye, and M. Gales. Iterative unsupervised adaptation using maximum likelihood linear regression. In *Fourth International Conference on Spoken Language Processing*, 1996.

[20] S. Young, D. Kershaw, J. Odell, V. Valtchev, and P. Woodland. *The HTK Book (for HTK Version 3.0)*. Microsoft Corporation, Redmond, WA, USA, 2000.

[21] H. Zhen, H. Wang, and N. Black. Human activity detection in smart home environment with self-adaptive neural networks. In *Proceedings of IEEE International Conference on Networking, Sensing, and Control*, pages 1505–1510, 2008.