Using Mobile Phones to Determine Transportation Modes

SASANK REDDY, MIN MUN, JEFF BURKE, DEBORAH ESTRIN,
MARK HANSEN, and MANI SRIVASTAVA
University of California, Los Angeles

As mobile phones advance in functionality and capability, they are being used for more than just communication. Increasingly, these devices are being employed as instruments for introspection into habits and situations of individuals and communities. Many of the applications enabled by this new use of mobile phones rely on contextual information. The focus of this work is on one dimension of context, the transportation mode of an individual when outside. We create a convenient (no specific position and orientation setting) classification system that uses a mobile phone with a built-in GPS receiver and an accelerometer. The transportation modes identified include whether an individual is stationary, walking, running, biking, or in motorized transport. The overall classification system consists of a decision tree followed by a first-order discrete Hidden Markov Model and achieves an accuracy level of 93.6% when tested on a dataset obtained from sixteen individuals.

13

Categories and Subject Descriptors: I.5.2 [**Pattern Recognition**]: Design Methodology—*classifier design and evaluation*

General Terms: Algorithm, Design, Experimentation

Additional Key Words and Phrases: Activity classification, mobile phones, transportation mode inference

ACM Reference Format:

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M. 2010. Using mobile phones to determine transportation modes. ACM Trans. Sensor Netw. 6, 2, Article 13 (February 2010), 27 pages. DOI = 10.1145/1689239.1689243 http://doi.acm.org/10.1145/1689239.1689243

1. INTRODUCTION

Mobile phones are truly ubiquitous. They have computation, sensing, and communication capabilities and are carried by people throughout the day. Many

This work was supported (in part) by National Science Foundation (NSF) grant number CNS-0627084 and Nokia Research.
This is an enhanced version of preliminary work that was presented at the International Symposium on Wearable Computers (ISWC 2008).
Authors' address: Center for Embedded Networked Sensing, UCLA, 3563 Boelter Hall, Los Angeles, CA 90095; email: {sasank, bobbymun, jburke, destrin, cocteau, mbs}@ucla.edu.

DOI 10.1145/1689239.1689243 http://doi.acm.org/10.1145/1689239.1689243

of these devices can already record audio, take photos, and communicate over different radio channels, and more recently these phones are being equipped with sensors that are able to capture location and measure acceleration. Taking advantage of these features on mobile phones, several practical systems have been implemented that have enabled applications in regards to sharing sensor derived status information in online social networks, capturing the characteristics and dynamics of everyday activities such as the commute, and enabling queries associated with physical space [Miluzzo et al. 2007; Gaonkar et al. 2008; Li et al. 2008; Mohan et al. 2008].

This work focuses on using mobile phones to determine the transportation mode of an individual when outside, whether the user is stationary, walking, running, biking or in motorized transport. Target applications for this fine-grained transportation mode inference and location information include the following.

—*Physical Activity Monitoring*. The transportation modes of individuals are logged and mapped to locations to enable individuals to plan modes of transportation based on goals of physical activity and for the purpose of health monitoring [Consolvo et al. 2008a; Peterson et al. 2009].

—*Personal Impact and/or Exposure Monitoring*. Inferences of the transportation mode and location of an individual are used to provide a personalized environmental scorecard for tracking the hazard exposure and environmental impact of one's activities. Examples include our Personal Environment Impact Report (PEIR) and UbiGreen [Mun et al. 2009b; Agapie et al. 2008; Froehlich et al. 2009] along with commercial offerings such as Ecorio and Carbon Diem [CarbonHero 2008; Kao et al. 2008].

—*Transportation and Mobility-Based Recruitment*. Transportation annotated mobility profiles (time, location, transportation mode traces) are created for profile based recruitment for distributed data gathering [Reddy et al. 2009b; Burke et al. 2006].

The accuracy requirements for these applications are high. In PEIR the allowable noise from the transportation mode classification cannot exceed 10% (accuracy of the classifier has to be greater than 90%). Analyzing several weeks of transportation mode activity of members in PEIR indicates that higher error rates compromise an individual's ability to make choices about their daily transportation habits; in effect adding noise to the impact and exposure estimates that is on par with "natural" variations that they may want to study (changes in speed or the selection of alternative routes). Similar high accuracy requirements are needed for the other two applications to avoid undermining user confidence in the system (in the case of the physical activity monitoring) and utility of data collection (in the case of transportation and mobility based recruitment). For instance, when monitoring physically activity for fitness, reporting instances of biking as motorized transport or walking significantly affects energy expenditure estimates for an individual and can cause feedback for goal setting to be incorrect. In the case of the recruitment problem, many of the data collections are related to certain types of transportation modes, such as monitoring the quality of cycling routes or documenting walking hazards [Mun et al. 2009a; Reddy et al. 2009a], so it is imperative that the mode classification is accurate otherwise the wrong set of individuals could be recruited for participation - thus reducing the utility of the data collection as a whole.

This article details the design, implementation, and evaluation of the transportation mode classification system that runs on a mobile phone equipped with a GPS receiver and a 3-axis accelerometer. The classification system is convenient for an individual yet reliable in accurately distinguishing between the five transportation modes. The fact that the classifier runs on a single sensing unit without strict orientation or position requirements makes the system convenient to use. Also, the unit does not rely on external indexes such as GIS information or historical user pattern data. The overall system achieves an accuracy of greater than 93% and works reliably even if user-specific training data is not present. The classifier is composed of a decision tree followed by a first-order discrete Hidden Markov Model and works by analyzing a second of GPS speed data along with variance and frequency components of the accelerometer signal. Since the transportation mode classification has an energy footprint associated with it, an algorithm was also created to turn on the classifier when an individual goes outdoors. This procedure relies on using changes in the connected cell tower as a trigger to check the outdoor status of an individual as opposed to uniformly sampling the GPS.

The remainder of the article is organized as follows. Related work is discussed in Section 2. Section 3 details the system design goals and contributions along with the sensor, feature, and classifier selection approach. The experimental setup is described in Section 4. Section 5 contains the classifier evaluation and performance results. Section 6 and 7 contain discussion of future work along with the conclusion.

2. RELATED WORK

Many systems exist to classify human motion activities and transportation modes. Related work in this space can be grouped based on the types of systems used to implement the algorithms: commercial devices, custom hardware, and mobile phones [Consolvo et al. 2008b]. Each category is detailed below with information on how our system builds on or differs from the ones described.

2.1 Commercial Devices

Devices from the commercial realm for activity monitoring vary in terms of the sensors used and inferences determined [Chen et al. 2008]. One of the most ubiquitous devices for physical activity monitoring is the pedometer [Consolvo et al. 2008b]. It consists of a sensor, such as a mechanical arm, magnetic switch, or an accelerometer, and software that counts steps based on monitoring upward and downward motions [Crouter et al. 2003; Schneider et al. 2004]. Pedometers are typically designed to be worn in a specific orientation (vertically) on the hip [Omron 2008a] or the ankle [Karabulut et al. 2005], but recent advances have made them more applicable to other positions and orientations on the body [Nike 2008; Omron 2008b]. Furthermore, mobile phones that have accelerometers have been used to keep track of step counts as well [Chen et al. 2008].

More sophisticated devices exist commercially for activity monitoring. For instance, both FitBit and Phillips Tracmor devices incorporate multi-axis accelerometers to provide a convenient (orientation agnostic) method to infer calories burned. Impact Sports' ePulse monitor uses a heart rate and BodyMedia's GoWear unit combines four sensors (accelerometer, heat flux, galvanic skin response, and skin temperature) for this same purpose [ImpactSports 2008; BodyMedia 2008]. Although these commercial offerings are widely available and fairly convenient, they only provide coarse activity information (step count, calories burned, distance traveled). Our applications require finer grained activity labels, such as which of the five transportation modes were taken, to be effective.

2.2 Custom Hardware

Research groups have investigated different methods to infer motion activities (climbing stairs, elevator rides) along with transportation modes using custom hardware Consolvo et al. [2008b]. For instance, Farringdon et al. [1999] and Randell and Muller [2000] have created systems that use a single accelerometer to infer stationary, walking, and running activities. The work from Kern et al. [2003], Bao and Intille [2004], Ganti et al. [2006], Fábián et al. [2008], and Saponas et al. [2008] has concentrated on using multiple accelerometers placed on different positions on the body to infer activities, and Pham and Abdelzaher [2008] have shown that orientation independence can be achieved in this multiple device setting. Unfortunately, the single accelerometer solutions cannot differentiate between being stationary and in motorized travel with high accuracy, and although multiple worn accelerometer solutions provide highly detailed information, they are only practical for specialized applications and use cases.

Similar to SenseWear, Lester et al. [2006] use a single sensing unit (Mobile Sensing Unit - MSP) with multiple modalities (accelerometer, audio, barometric pressure) for activity inference. They focus on user convenience and show that their system works across several users and at multiple positions. In more recent implementations, audio is eliminated as a sensor modality, and instead either identifier/signal strength information from multiple cell towers (up to seven) [Froehlich et al. 2009; Sohn et al. 2006] or coarse speed measures from network endpoint localization [Welbourne et al. 2005] is used for transportation mode inference. But unfortunately, obtaining multiple cell tower information for Windows Mobile 6, Nokia Symbian, Android, and the iPhone based phones is not possible using the standard developer application programming interfaces. These platforms limit the cell information to only the connected tower or do not make it available at all. Furthermore, the WiFi and GSM fingerprinting needed for network endpoint localization is limited to predominantly urban areas and only provides a rough estimate of speed [Skyhook 2009; Wigle 2009; LaMarca et al. 2005]. Finally, the MSP device used has a predetermined fixed orientation and attachment procedure associated with it.

Table I. Related Work Implemented on Mobile Phones

	Classes	Sensor	Users	Time	Accuracy
Anderson and Muller [2006]	Still, Walk, Motorized	GSM	1	45 Mins.	82%
[Sohn et al. 2006]	Still, Walk, Motorized	GSM	3	323 Hours	85%
[Mun et al. 2008]	Still, Walk, Motorized	GSM, WiFi	2	13 Hours	83%
[Liao et al. 2007]	Walk, Motorized (Bus/Car)	GPS, GIS	1	60 Days	84%
[Zheng et al. 2008]	Walk, Bike, Motorized	GPS	65	10 Months	76%
[Miluzzo et al. 2008]	Still, Walk, Run	Accelerometer	8	4 Hours	78%

Our work expands on these existing systems by enabling fine-grained transportation mode classification based on sensors (GPS and accelerometer) available on a commodity device (mobile phone) along with relaxed requirements on how the device should be worn (any orientation) in addition to being position (6 different) and user agnostic. We note that the approach of using GPS and accelerometer sensors for transportation mode inference was first mentioned by Marmasse et al. [2004] but never throughly evaluated, and that Denning et al. [2009] and Lester et al. [2008] consider this method as a future direction for applications that use the MSP.

2.3 Mobile Phones

With the advancement of sensors on mobile phones, researchers and practitioners are looking to use this device as a platform for activity inference [Consolvo et al. 2008b]. Table I shows a summary of work that has taken place in this space along with the types of activity modes inferred, the test user base, and the classification accuracy. Anderson and Muller [2006] and Sohn et al. [2006] use changes in GSM cell tower observations (up to seven) to approximate whether a user is still, walking, or in motorized transport. Mun et al. [2008] augments the GSM work (but uses only the connected cell tower) with the addition of features derived from WiFi observations for classification purposes. Both GSM and WiFi based systems work well for coarse grained transportation mode classification, such as determining the difference between still, human powered motion, and motorized transport, but are not as useful for fine-grained classification. Also, these features are dependent on network endpoint density which varies based on the environment the user is in.

Several systems rely on GPS combined with external information for sensing transportation modes. For instance, Zheng et al. [2008] combine GPS with a post-processing step that uses likely transportation modes from a corpus of contributed data for classification. Alternatively, Patterson et al. [2003] combines GPS with GIS information, such as transportation end points and road networks, and Liao et al. [2007] builds user specific models based on learning destinations and routes from historical data to build transportation mode classifiers.

Our work differs from existing solutions in that it does not rely on external indexes since this data might not always be available. Also, GPS only solutions work well for coarse grained transportation mode classification, but perform poorly when classification of travel modes with similar speed and acceleration profiles is needed, such as the case with running, biking, and slow motorized travel. Miluzzo et al. [2008] uses a mobile phone with a three-axis accelerometer for inferring different classes of walking motion. We show later that both GPS and accelerometer features are needed when considering additional transportation modes.

3. APPROACH

This section contains details in regards to the underlying design principles and the core contributions associated with the transportation mode classifier. Furthermore, information about the sensor modalities and classifier types considered for our system is provided.

3.1 Design Goals and Contribution

The primary design goal for the transportation mode classifier is user convenience. Thus, our system has the following properties:

—contained in one sensing unit;
—flexible in terms of the position and orientation;
—able to work for a variety of users without additional training;
—effective with sensors that exist on mobile phones;
—not reliant on external spatial or user history based indexes.

Although there exists systems for transportation mode classification, very few meet our design requirements. Many rely on having multiple devices placed at different positions on the body, having set orientation/position requirements, accessing external spatial or historical user pattern information, or using sensing modalities that are not viable or available to be used on mobile phones. The systems that are convenient for users and run on mobile phones are not as reliable for fine-grained transportation mode classification since they use a single sensing modality (accelerometer, GPS, or network endpoints). Thus, our contribution is a classifier that uses information from an accelerometer and a GPS to achieve high accuracy and is able to run on a commodity mobile device that varies in terms of how it is worn, carried, and used. Furthermore, we thoroughly evaluate the system by showing the usefulness of the sensing modalities employed, justifying the specific classification algorithm chosen, and illustrating that the design goal of user convenience is met through a series of tests based on annotated data collections. Finally, since our applications need to determine the transportation modes along with location when an individual is outside, we implement a energy-aware scheme that automatically turns on the transportation mode classifier during outdoor operation by using GSM cell tower changes as a triggering mechanism.

3.2 Sensor Selection

Since the classifier is designed to be run on a mobile phone system, there are a few different sensor modalities available for classification. In addition to accelerometer and GPS capabilities, modern phones are being equipped with Bluetooth, GSM, and WiFi radios. Features can be derived to estimate speed based on the existence and signal strength of Bluetooth beacons, GSM cell towers or WiFi access points. Below we discuss the different sensing modalities in more detail and provide justification for choosing GPS and accelerometer for the final classification system. Although our decision in regards to the sensing modalities to use could be guided by relying on previous work, we re-evaluated the sensing choices here since we are considering a specific set of fine-grained transportation modes.

3.2.1 *Bluetooth.* One option is to use Bluetooth as a sensor for activity classification. Previously, static Bluetooth beacons distributed throughout an indoor setting have been employed for determining activities based on proximity to devices (watching TV, washing devices, cooking), but this sensing modality has not been used for inferring transportation mode classifications [Tapia et al. 2004]. The main reason that Bluetooth has not been applied to transportation mode classification is because Bluetooth sensors are not ubiquitous in outdoor settings. Static Bluetooth beacons mainly exist in indoor settings such as office buildings or homes. Relying on Bluetooth signals exhibited by mobile phones has its own problems. For instance, it is difficult to distinguish whether an individual is moving or if the environment around them is changing (other people carrying devices are moving). Hence, Bluetooth is not an effective modality for transportation mode classification since it does not provide the accuracy needed for the modes being distinguished.

3.2.2 *WiFi and GSM.* To determine whether WiFi and GSM features would be useful for the fine-grained transportation mode classification, a test using training data consisting of collecting accelerometer, GPS, WiFi, and GSM raw data from sixteen individuals in an urban settings was performed. More information about the setup of the data collection can be found in Section 4.3, and the classifier used for this preliminary sensing modality analysis was a C4.5 Decision Tree. When comparing WiFi and GSM features (note that only the connected cell information was used) [Mun et al. 2008] to accelerometer and GPS features (further described in Section 3.3), the results indicate that the accelerometer and GPS features enabled higher accuracy classifiers - around 22% greater with an overall accuracy of 91.3%. Furthermore, classifiers built using accelerometer information with WiFi, GSM, and GSM/WiFi features, still resulted in the GPS and accelerometer based classifier to be 3–7% greater in accuracy. Table II shows the drop in accuracy when using features from different modalities as compared to the accelerometer and GPS based classifier. Each row shows which modalities were employed in creating the specific classifier being compared. Overall, the results presented in the table make sense for two reasons: (1) GSM and WiFi features act as a proxy for the speed of an individual and when speed profiles are similar, as is the case with slow moving traffic, biking, and walking, these features are not as discriminative; and (2) GSM and WiFi features depend on the density of network end points (cell towers and access points) and even in urban settings, there are areas (recreational, low residential) where the density is low or the end points cover a large region.

Table II. Classification Accuracy Decrease Compared to GPS and Accelerometer Based System

Accelerometer	GSM	GPS	WiFi	Accuracy Decrease
X				10.4%
	X			33.2%
		X		19.2%
			X	35.1%
X	X			6.9%
X			X	3.9%
X	X		X	3.0%
	X	X		13.3%
	X		X	22.1%
		X	X	11.9%

We also analyzed whether all modalities should be used for a classifier or if we could simply use accelerometer and GPS features. The classifier with all four modalities resulted in a negligible increase in accuracy (0.6%) as compared to using just GPS and accelerometer features. Thus, we conclude that GPS and accelerometer features are effective for our transportation mode classification system, and that using WiFi and GSM is not necessary since they add minimal performance increase while significantly impacting the energy footprint of the classifier (see Section 5.6 in regards to the power usage of sampling these additional modalities). Furthermore, using GPS enables the system to obtain granular location information which is useful for our target applications, and not relying on WiFi and GSM features avoids having to create environment specific models to account for network endpoint characteristics in urban, suburban, and rural areas [Mun et al. 2008; Chen et al. 2006; Sohn et al. 2006; Anderson and Muller 2006].

3.2.3 *Accelerometer and GPS*. Accelerometer and GPS information are both needed for transportation mode classification. In situations where the accelerometer output is similar, the speed is typically different and vice versa. If the system employs just one of these sensors for classification, a drop in accuracy of 10–20% results when compared to using both. Transportation modes such as still, biking, and motorized travel with accelerometers and walk, run, and bike for the GPS were the most affected since they had similar profiles when only one sensor is used.

The usefulness of both modalities is further shown in Figure 1. Figure 1(a) contains the speed (meters per second) distribution of 30 minutes of each activity sampled every second from one individual. We note that the exact makeup of the speed distributions can differ on a per user basis (individual's could have a faster/slower walking stride or have different driving habits) and urban canyons can cause speed outliers, but the figure illustrates the "ranges" of speeds for the transportation modes, which were consistent among all individuals in the user base. For instance, stationary activities have speeds mainly under 1 meters per second and walking typically has speeds of 0.5–2 meters per second. Running and biking are similar in their speed distribution while motorized activity has a larger range that is usually higher then the other modes. Accelerometer data is very useful when the activity causes a change in motion.

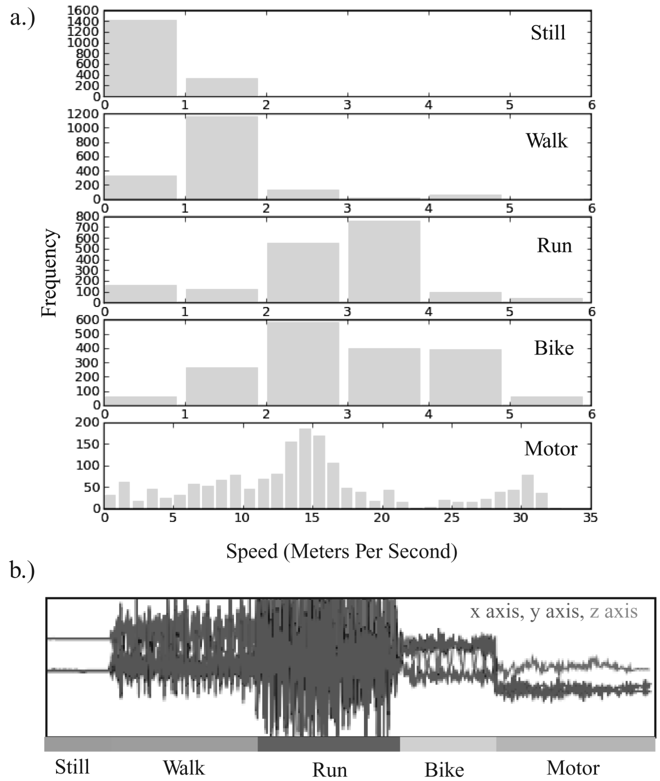

Fig. 1. Distribution of speeds along with variations in accelerations for the transportation modes.

For instance, stationary activities have a very low variance. Motorized transport is similar to stationary activities but is affected by vehicle vibrations and road conditions. Walking and biking exhibit similar accelerometer characteristics in certain areas of the body, and running has a large variance. Figure 1(b) shows accelerometer data from all three axes (different colors - red, green, blue) for the case where the user is still initially for a short period of time, then starts walking, running, biking, and is in motorized travel.

3.3 Feature Selection

This section details the decisions involved in selecting the features for the classification system. First, the methodology for picking the raw sample window size to make into features is reviewed. Then, the types of features that were considered are detailed. Finally, information about which features were chosen for the classification system along with the criteria for the decision making process is provided.

3.3.1 *Window Size.* A window of 1 second is used for the period of classification. This value is validated by previous work in classifying the classes of activities this work targets [Bao and Intille 2004; Huynh and Schiele 2005;

Kern et al. 2003]. Smaller window sizes cause classification accuracy to suffer due to certain features (accelerometer frequencies) not being effective and larger window sizes introduce noise since multiple activities could exist. Furthermore, a larger window was not chosen so that the classifier would work immediately when instigated.

3.3.2 *Types of Features.* The two sensors that are sampled for data include the accelerometer and the GPS. In terms of the accelerometer, we use the magnitude of the force vector by combining the measurements from all three axis as the basis for the features: $A_{mag} = \sqrt{(A_x)^2 + (A_y)^2 + (A_z)^2}$. This enables our system to assume a random and possibly changing orientation for the mobile phone. Various features including the mean, variance, energy, and the DFT (Discrete Fourier Transform) energy coefficients between 1-10Hz based on magnitude of the force vector of the accelerometer were evaluated along with the speed of the GPS receiver [Bao and Intille 2004; Kern et al. 2003]. Frequencies between 1-10Hz for the accelerometer were chosen for consideration since prior work indicates this range is appropriate to detect pedestrian based motion [Welbourne et al. 2005; Sun and Hill 1993; Winter 1990; Antonsson and Mann 1985]. In terms of speed, the value obtained from the GPS receiver directly is used, if available, since it is more accurate than calculating speed from consecutive location points.

Note that there is a noise filtering step that occurs for the classification system. GPS points deemed “invalid”, based on analyzing the accuracy (vertical, horizontal, heading, and speed), dilution of precision (time, vertical, horizontal), and changes in speed values of the signal, are discarded. These invalid points normally occur when the phone is significantly shielded or if the person is in an area that is covered. The filtering process also analyzes accelerometer data. If too few samples are received from the accelerometer to calculate the frequencies of interest, this data is excluded for classification as well.

3.3.3 *Selection Method.* Variance along with DFT energy coefficients between 1-3 Hz from the accelerometer and the speed from the GPS receiver were selected as the feature set using correlation based feature selection (CFS). CFS was chosen as opposed to Principle Component Analysis (PCA) because it is a feature subset selector that eliminates irrelevant and redundant attributes. Essentially, CFS uses a heuristic “merit” function that finds the subset that is predictive of the classification groups while reducing redundancy among the features themselves [Ghiselli 1964; Yu and Liu 2003]. On the other hand, PCA is concerned with creating new (and usually fewer) feature directions from linear combinations of original features and transforming data samples onto the new feature space [Martinez and Kak 2001]. This is an expensive operation since it requires all the original features to be calculated and then transformed into the new projection space. Furthermore, using PCA, even with 95% of the variance explained, results in a 2-3% less accurate classifier than just relying on CFS based subset features.

To show the “worth” of each feature in the subset, the information gain (or the predictive power) is measured of each attribute. As Table III shows, the GPS

Table III. Information Gain Scores of Features Involved in Subset Chosen

Feature	Score
GPS Speed	1.431
Accelerometer Variance	1.426
Accelerometer DFT (3 Hz)	1.205
Accelerometer DFT (2 Hz)	1.125
Accelerometer DFT (1 Hz)	0.915

speed value is the most important with the highest information gain score. The variance of the accelerometer signal is close in terms of value and then come the three DFT energy coefficients from the accelerometer. This result makes sense since the GPS speed can help to determine when a user is still or in motorized transport, the accelerometer variance can be used to infer if an individual is running, and the DFT coefficients help in differentiating between the foot-based transportation modes.

3.4 Classifiers Selection

To determine which classification system is the most accurate for transportation mode inference, we compared: (a) instance classifiers such as C4.5 Decision Trees (DT), K-Means Clustering (KMC), Naive Bayes (NB), Nearest Neighbor (NN), and Support Vector Machines (SVM), (b) a continuous Hidden Markov Model (CHMM), and (c) a two-stage system involving the most accurate instance based classifier (which is the decision tree) combined with a discrete Hidden Markov Model (DHMM).

3.4.1 *Classifier Details.* We point readers to Duda et al. [2000] and Witten and Frank [2005] for definitions of the commonly known classifiers but go into more detail in regards to the CHMM and the two-staged classification system which follows for clarity:

—*Continuous HMM (CHMM)*. CHMM is a hidden markov model where the output symbols (features) are modeled as independent multi-variate Gaussian distributions and the hidden states correspond to classification classes Also, the transition probability between classes is also considered.

—*Instance Classifier + Discrete Hidden Markov Model (DHMM)*. The two-stage classifier is a instance based classifier followed by a DHMM where the DHMM is trained by the class posterior probability of the instance based classifier. Thus, the DHMM output symbols are the instance-based classifications and the hidden states are the classification classes. Similar to CHMM, the state transition probabilities are also modeled.

3.4.2 *Parameters and Specification.* Many of the classifiers require selection of thresholds and parameters. We review our decision making process to select these attributes for the classifiers below. In terms of DTs, a major concern is over-fitting, and thus both pre-pruning and post-pruning are employed. For pre-pruning, the minimum number of points needed for each leaf needs to be set, and for post-pruning, the reduced error technique is employed [Esposito et al. 1997]. The KMC classifier requires the number of clusters to be chosen a priori. Setting the cluster number to 5, which is also the number of classes in the system, results in the most accurate classifier.

In the NN model, the number of closest neighbors to consider is 1. Higher order nearest neighbor classifiers are not used since they do not provide significant increase in classification accuracy and require more distance instances to be maintained for classification. For the SVM classifier, a linear kernel was chosen as the basis for the hyperplanes due to its short training time and feature transformation computation simplicity. Since the system has multiple classes to consider, pairwise binary classifiers are created for the different states, and classification is performed by finding the class with the highest number of votes for the input feature space.

The CHMM has five hidden states corresponding to the transportation modes and the output symbols are the accelerometer and GPS receiver features modeled as independent multi-variate Gaussian distributions. The two-stage system uses an instance based classifier followed up by a DHMM where the instance based classifier directly uses the raw features and the DHMM is trained by the class posterior probabilities of the instance based classifier. Thus, the DHMM output symbols are the initial classifications and the hidden states are the transportation modes. The state transition probabilities are set for the HMMs based on eliminating transitions that are unlikely to happen, such as going from biking to running or motorized transport to biking. The specific values were chosen empirically by testing on labeled transportation mode traces. Alternative approaches for modeling the HMM structure include using fewer or more hidden states, however this would result in hybrid modes that would have to be reinterpreted back to the base set of five transportation modes that are desired as output for classification. Furthermore, the classifier could consist of several HMMs corresponding to each of the different transportation modes with the features being used to figure out which HMM a sequence best matches, but this structure would not provide a method to smooth transportation mode inferences based on transition probabilities.

4. EXPERIMENTAL SETUP

This section details the hardware platform, software setup, and the data collection involved in creating the transportation mode classifier. Specifically, we provide information on the type of cell phone used along with the exact software setup involved in the training, testing, and final implementation for the system.

4.1 Hardware Platform

The system is implemented on the Nokia n95 [Nokia 2008]. This device was chosen due to its sensing functionality and form factor. Along with its 332 MHz ARM processor and 128 MB of RAM, it contains a three axis accelerometer with a sensitivity of +-2G and that can sample at 32 Hz and a built-in GPS receiver that can sample at 1 Hz. Furthermore, a WiFi radio that can scan at 0.33 Hz, GSM cell radio that can sample at 1 Hz, and a Bluetooth radio that can scan at .08Hz are included in the device as well. The capacity of included battery is 950 mAH.

4.2 Software Setup

To evaluate different classification schemes, the Weka Machine Learning Toolkit and the Generalized Hidden Markov Model library were employed [Witten and Frank 2005; GHMM 2008]. The final chosen classifier is run on the Nokia n95 and programmed using Python for Symbian S60. Python was chosen since it enables rapid development, porting to other platforms, and does not have code signing restrictions that require user involvement for the sensing operations.

4.3 Data Collection

The data set used for training and testing of the transportation mode classifier was obtained by asking sixteen individuals, eight male and eight female between the ages of 20-45, to gather fifteen minutes of data while outside for each of the five transportation modes. The volunteers performed the activities in an urban setting with six phones attached simultaneously — positioned on the arm, waist, chest, hand, pocket, and in a bag with orientations set according to their preference. Accelerometer, GPS, WiFi, and GSM information were obtained according to the sample rates described earlier. In order to have GPS speed information available throughout the data collection, a GPS lock was obtained initially and the participants were advised to keep the keypad of the phone in the exposed position (slid open). In general, keeping the keypad in the exposed position (as instructed by the Nokia n95 manual) enabled us to maintain a consistent GPS lock even when the phone is covered by clothing or placed inside a bag. The participants had a choice of using a back-pack, fanny-pack, or a tote (large open purse).

Instructions were given as to the duration of each activity needed and participants were advised to represent different styles with which they would perform each activity. The volunteers concentrated on one transportation mode at a time and performed all five consecutively during their data collection session. Ground truth annotations were controlled by the individuals, and post filtering was performed to eliminate ambiguous states (being stationary on a bike or in motorized transport). The total amount of data collected across all sixteen individuals was 120 hours, compromised of 1.25 hours of data per position (six) per individual (sixteen).

In addition to the collection described above, two additional data gathering efforts were performed. The second data collection involved one volunteer (who was involved in the primary collection) running the classification system while annotating transportation modes during everyday operation in typical and challenged environments. More information in regards to this data collection is provided in Section 5.4. The third data collection involved sixteen individuals annotating their full day (on average 23.2 hours with a minimum of 20.7 hours and a maximum of 26.8 hours) in terms of transportation modes and indoor/outdoor status while collecting the GSM cell tower identifier (1 Hz).

Table IV. Precision Results for Classifiers

	Still	Walk	Run	Bike	Motor	All
DT	**95.0**	**87.6**	**95.5**	**84.5**	**93.9**	**91.3**
KMC	54.0	81.0	98.5	45.6	98.9	75.6
NB	88.4	88.1	93.5	75.6	71.3	83.4
NN	96.4	87.3	93.3	84.8	92.7	90.9
SVM	90.7	88.8	95.9	81.6	97.8	91.0
CHMM	89.2	90.0	94.3	80.5	77.6	86.3
DT-DHMM	**95.5**	**92.4**	**96.4**	**87.9**	**96.2**	**93.7**

Table V. Recall Results for Classifiers

	Still	Walk	Run	Bike	Motor	All
DT	**97.2**	**88.4**	**91.9**	**85.3**	**93.4**	**91.3**
KMC	99.7	75.3	81.0	34.8	63.2	70.8
NB	97.2	77.4	94.2	51.2	95.3	83.0
NN	96.6	88.0	92.9	84.2	92.9	90.9
SVM	97.4	86.9	92.7	87.1	89.4	90.7
CHMM	97.5	79.0	94.7	63.5	95.9	86.1
DT-DHMM	**97.8**	**90.8**	**94.4**	**90.6**	**94.5**	**93.6**

The annotated days consisted of 8 weekday and 8 weekend periods. This dataset is used to evaluate our algorithm for turning on transportation mode classification only when an individual is outside.

5. RESULTS

In order to analyze the performance of the transportation mode classifier, three distinct metrics are employed: accuracy, precision, and recall [Compumine 2008]. The testing setup has even numbers of every class so the overall accuracy is simply the average of the recall values for each of the transportation modes [Witten and Frank 2005]. The first test below, which shows a comparison of different types of classification models, uses the precision and recall attributes to measure performance. Having both metrics enables an intuitive feel for the behavior of the individual classification techniques. For the placement and user based tests, the accuracy measure is only used since the precision and recall values are similar.

5.1 Classification Accuracy

To test the different instance based classifiers, 10-fold cross validation is employed where the folds contain equal amounts of each activity and are made up off random continuous segments from the experiment data set. Table IV and V show the precision and recall values for the classifiers. In terms of the instance based classifiers, the DT is the most effective with an overall precision and recall levels both equal to 91.3%. The DT had 31 nodes, 16 leaves, and a depth of 7 levels. When the DT was combined with a DHMM, the precision and recall improve to 93.7% and 93.6%. Thus, our final classification system, Figure 2, is made up of these two classifier modules combined together. The output of the classification system is the transportation mode with the associated time and GPS location information.

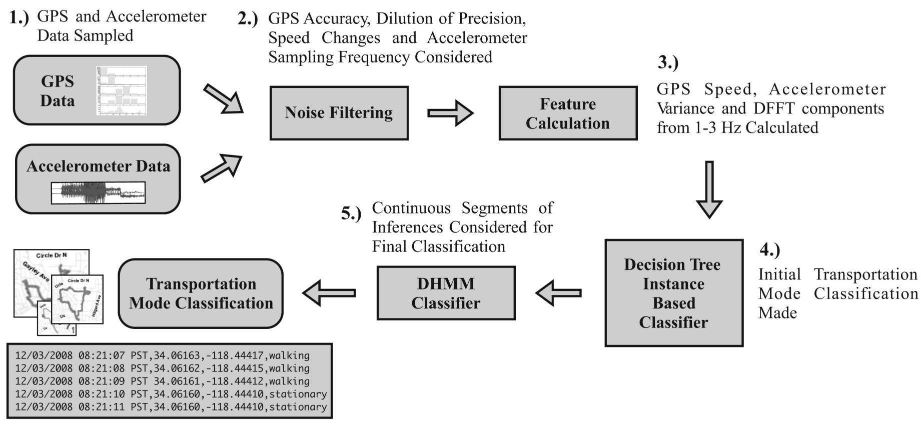

Fig. 2. Structure of Overall Classifier.

It is plausible that the DT and DHMM combination is the most accurate (combining the other instance based classifiers with the DHMM results in lower accuracy) since the DT is tuned to differentiate between the boundaries of transportation modes, and the DHMM eliminates noise based on temporal knowledge of the previous transportation mode and the likelihood of transitioning into the next mode. The DHMM is especially useful in helping fix problems with the non-stationary states. For instance, there are cases where being on a bike exhibits similar feature characteristics as running (especially if the individual is peddling fast) but since its unlikely that there will be transition from biking to running this inference is corrected. The same idea applies for the other states that were improved by the two-stage classification system. Our classification system is similar to the one employed by Lester et al. [2005] in that discriminative (DT) and a generative (DHMM) models are used in conjunction [Jaakkola and Haussler 1999]. But our work differs in that our system uses a single DT and DHMM pair that models likely transitions between transportation modes instead of an ensemble of classifiers feeding into several HMMs that represent each class. The single DHMM with all transportation modes and transition probabilities modeled was the most effective structure for type of errors (outlined above) that were encountered with the DT.

Details in regards to why the other classifiers do not perform as well or are not employed in the final classification system are provided below. The KMC algorithm has lower accuracy levels since the features do not naturally cluster into separable class spaces needed for classification (especially with the biking state). The NB algorithm effectiveness suffers since the probabilistic independencies among the features does not hold - speed, variance, and DFT energy coefficients (1-3Hz) are related to certain extent in terms of the classes. For the CHMM, the fact that in the system several of the classes have features, or outputs in CHMM terminology, that overlap in terms of their multi-variate Gaussians makes the accuracy of the classifier to be lower. The SVM classifier has high accuracy but its performance is still lower than the DT. The NN

Table VI. Phone Position and Accuracy

	DT+HMM
All Positions	93.6
Arm	94.9
Bag	94.8
Chest	94.5
Hand	95.0
Pocket	94.3
Waist	94.4

method is specifically not preferred due to its computation footprint since distance calculations between every instance in the training set needs to be made for an inference to take place.

5.2 Device Placement Variation

Given the goal of user convenience, we investigated how phone placement affects transportation mode accuracy [Bao and Intille 2004; Lester et al. 2006]. Mobile phones are often carried by individuals at different positions, and requiring a set position is inconvenient. For instance, males often attach their mobile phones to a belt holder and carry them in their pocket; females often put their phones in a bag; and individuals that are performing exercise often attach the phone to the arm or chest area. For testing purposes, a general DT+DHMM that is trained on data from all six positions (arm, bag, chest, hand, pocket, waist) and individual DT+DHMMs trained on data from specific positions were created using 10 fold cross validation.

The results, shown in Table VI, indicate that a generalized classifier is on par in performance to ones trained on specific positions. The average accuracy decrease for the generalized classifier is only 1.1%. Thus, a generalized classifier can be created so that the user can be agnostic about where to position the phone and still obtain accurate transportation mode inferences. In terms of the position specific results, the pocket, waist, and chest have lowest accuracy levels since these can be impacted by body motions such as bending, swaying and twitching. The arm and hand positions were the most accurate. Transportation modes such as walking, biking, and running have "strides" that show up at these positions. Placing the phone in a bag enables high inference accuracy since this position has less variability in movement and is near the body.

The generalized classifier has additional features that make it convenient for users. For instance, a set orientation is not required for the phone since the force vector is used for the basis of the accelerometer features and speed of the GPS is employed. Furthermore, the device can be placed inside typical clothes and bags or exposed externally. None the less, the system does have some drawbacks. Currently, the Nokia n95 keypad needs to be in the exposed position in order to obtain a GPS lock quickly and maintain it in a consistent fashion. This phone configuration is the recommended operation mode for our transportation mode classifier. However, if the phone is in a closed state, our experiments show that the system still performs well by sacrificing quantity of outputs to

Table VII. DT+DHMM User Specific Classification Accuracy Results

User 1	94.5	User 9	96.0
User 2	97.9	User 10	97.6
User 3	93.3	User 11	93.4
User 4	93.8	User 12	96.6
User 5	92.7	User 13	96.3
User 6	96.8	User 14	96.8
User 7	98.0	User 15	95.2
User 8	97.1	User 16	96.7
		Average	95.8

maintain performance - filtering occurs to get rid of lower quality GPS signals by analyzing dilution of precision, accuracy, and changes in speed values, and in general GPS locks are often lost or not maintained in this phone configuration. It is important to note that this preference of leaving the keypad in the exposed position is device dependent since the GPS quality issue does not exist for the iPhone or TMobile G1 platform where preliminary tests have indicated that the GPS works well without any restrictions on how the device should be configured. Finally, if the phone is placed in a location that is significantly shielded, the device is not able to obtain GPS speed values and does not function. These types of situations include putting the device significantly distant from windows in motorized transport or covering the phone with a very thick material. As future work, we plan to explore a system that can intelligently back-off to using less accurate speed estimates derived from network endpoint based features/localization techniques when GPS is not available [Mun et al. 2008; Welbourne et al. 2005].

5.3 User Variation

Another factor related to user convenience is whether a generalized classifier can be created that would work for new users without individual user-specific training [Bao and Intille 2004; Lester et al. 2006]. To test the feasibility of such a system, two distinct experiments are performed with the DT+DHMM combination: (1) user specific mode where only a particular user's data is used for training and testing purposes with 10-fold cross validation and (2) leave one user out mode where the classifier is trained with all but one user (fifteen out of sixteen) and tested with the user not in the training set.

Table VII shows results from the user specific mode testing. When training and testing is done on an individual user basis, the accuracy increases by 2.2% compared to a generalized classifier that is trained and tested on all individuals. Thus, creating user specific classifiers would help in terms of classifier performance, although the increase in accuracy is minimal when compared to using a generalized classifier. With leave one user out mode, an average accuracy of 93.6% and a minimum accuracy of 88.2% is obtained. Table VIII shows the results for all sixteen users. Based on the results, one can conclude that certain users might be unique and a training set that has a broad range of how activities could be performed is necessary. In fact, for the user's that had the worst performance in terms of accuracy (user's #3, #5, #11, and #15), the

Table VIII. DT+DHMM Leave One User Out Accuracy Results

User 1	93.0	User 9	93.1
User 2	93.4	User 10	96.8
User 3	88.2	User 11	91.6
User 4	92.9	User 12	95.7
User 5	90.9	User 13	92.2
User 6	96.3	User 14	95.2
User 7	95.6	User 15	91.0
User 8	96.4	User 16	95.7
		Average	93.6

decrease in performance mainly came in the walking, running and biking activities for which individuals often have different styles both in terms of intensity and speed.

Overall, the results of two experiments indicate that it is possible to achieve good performance without requiring users to provide specific training data as long as the training set contains enough variation in terms of each activity. Furthermore, obtaining this variation could be done with few participants in the training set and may not require a large mass of individuals. Even with fifteen individuals, the minimum accuracy level was still above 88%. However, testing involving a larger population is needed to validate the hypothesis made based on this specific study.

5.4 Extended Transportation Mode Traces

Thus far, the classifier has been evaluated based on the explicit data collection performed by our base set of sixteen users. But we were interested in determining how the DT+DHMM classifier would perform in "everyday" use. Thus, we asked one of the individuals involved in our initial study to carry the mobile phone running the classifier over a period of four weeks and to document instances of each of the transportation modes when possible. The idea was to obtain longer traces of each of the modes to test the classifier's performance and to learn about scenarios where the classifier was not as effective, which we could only do by using the classifier in everyday situations. Since documenting ground truth labels for the transportation modes is an inconvenient and sometimes even a dangerous process, we allowed the volunteer to not have to document instances of being still when annotating modes that were not stationary (i.e., being still at a red light while in motorized transport, waiting at an intersection when walking or biking, or resting during sessions of running). We note that for our classifier these ambiguous instances will be outputted as still. This is done so that applications using the transportation mode classifier can handle these situations based on their specific requirements (i.e., a system to monitor physical activity would handle these cases differently from an impact/exposure measuring service). The participant had the liberty of carrying the device in any orientation and position that was desired, but instructions were given to have the phone keypad in the exposed position whenever possible in order to obtain and maintain a GPS lock. The results from the extended transportation mode traces are presented in Table IX and show clearly that

Table IX. DT+DHMM Accuracy for Extended Transportation Mode Traces

Transportation Mode	Length (Minutes)	Accuracy
Still	236	95.6%
Walk	227	96.8%
Run	103	91.0%
Bike	156	92.8%
Motor	345	93.9%

the DT+DHMM classifier is effective even in longer term everyday use. The accuracy (on average 94.0%) is similar to the performance that was found when analyzing our base set of sixteen users.

The transportation modes that had lower performance were motorized transport, biking, and running. For the case of motorized transport, the errors were introduced during slow acceleration after being stopped in traffic. These motorized transport instances were mistaken as biking since coasting on a bike (not peddling) is similar in terms of accelerometer and speed information. Also, the DHMM was not as effective in fixing these problems since slow travel occured for an extended period of time. The errors associated with biking and running came from misclassification as walking. In terms of biking, these errors occurred when the participant initially started biking after stopping for a period of time, and for running, the majority came up during slower speeds where the state could be considered either running or walking. We are exploring smoothing techniques, such as considering a window of transportation mode activity instead of just the previous classification for the DHMM, to help fix these particular problems.

In addition to analyzing the classifier performance in extended traces, we were interested in the accuracy when the system was intentionally used in challenging urban environments. Thus, we advised our volunteer to perform data collection in urban canyons (streets that are in between blocks of large building structures). Overall, close to 3.5 hours of data was gathered in this specific environment with each transportation mode consisting of at least 30 minutes. The resulting classifier had an average accuracy of 92.6%. Still, walking, and motorized transport performed the best with over 95% accuracy, while both biking and running states had lower accuracies of around 88%. Surprisingly the motorized transport state had a high accuracy, but from analyzing the traces this could be due to the fact that the travel that occurred did not involve extended periods of slow traffic. The biking and running errors were again due to mistaken outputs of walking. The path errors caused by the urban canyons seem to affect the lower end speeds more. In general the classifier performed at an adequate level, but improvements can be made by incorporating more challenged environment ground truth in the base training set.

5.5 Memory and CPU Benchmarks

In addition to the generalizability of the classifiers, the amount of resources in terms of CPU and memory (RAM) that the transportation mode classifier takes up while running on the Nokia n95 is of interest. Using the Nokia Energy Profiler, twenty minute trials were performed (where the individual metrics were

Table X. Memory and CPU Usage for Various Applications

Activity	CPU %	RAM (MB)
Phone Idle	2.18	28.91
Active Call	2.31	30.00
Music Player	30.86	30.26
Video Player	14.63	32.58
Game Playing	97.34	37.52
Transport Mode	**6.91**	**29.64**

Table XI. Power to Operate Various Sensor Types, Communication Methods, and Applications

Activity	Power (Watts)
Phone Idle	0.054
Cell Sampling	0.056
Accelerometer Sampling	0.111
Bluetooth Sampling	0.233
GPS Assisted Lock	0.718
GPS Normal Lock	0.407
GPS Sampling	0.380
WiFi Sampling	0.230
Music Player	0.447
Video Player	0.747
Active Call	0.603
Gaming	1.173
Transport Mode	**0.425**

averaged over the time span of the tests) that compared the resources used by the transportation mode classifier to other activities that normally occur on the phone. Table X shows the results. The CPU resources used represents the percentage of overall CPU taken up and the memory usage shows the amount of RAM in megabytes used. The transportation mode classifier uses CPU resources (6.91%) far below optional services, such as playing games, music, or videos, and the memory usage of the classifier is just above the phone being idle (29.64 megabytes). Thus, running the classifier will not significantly impact the performance of other applications that a mobile phone user might run. The feature extraction and DT+DHMM classifier takes up 1.85% of the CPU while the sampling makes up 2.88%. The sampling CPU usage is due to the sensor callbacks being setup and called. For instance, the accelerometer callback gets initiated roughly 32 times per second. The feature calculations and classifier output only get processed every one second. Thus, the processing involved for the classifier is not significant.

5.6 Energy Consumption

The energy consumption of the classifier is important. Table XI contains the energy usage of the transportation mode classifier along with information in regards to sampling various sensors and performing tasks on the mobile phone.

The results were obtained by analyzing 5 - twenty-minute trials for each of the activities. Furthermore, the screen was off for all the tests except for video playback and game execution which require the screen to be active to run. The GPS and GSM sensors were sampled at 1 Hz, WiFi at 0.33Hz, and the accelerometer reported readings on average at 32 Hz. The BT radio scans occurred every 10-15 seconds since this operation takes a longer time to complete.

The transportation mode classifier power usage (0.425 watts) is on par with playing music (0.447) while far below taking phone calls (0.603), watching videos (0.747), or playing video games (1.173). The power usage for processing features and running the classifier is relatively small (0.003). Thus, most of power required for the system is used up for sampling. Note that the Goertzel algorithm is employed to calculate the frequency components since only certain frequencies are desired [Welbourne et al. 2005; Goertzel 1958]. Running only the classifier on a typical Nokia n95 battery (950mAH) in an outdoor setting would result in 8.27 hours of operation. Overall, the transportation mode classifier uses significantly less energy then “add-on” entertainment and voice services.

In calculating the energy footprint of the transportation mode classifier, the phone was placed in assisted GPS mode. But even with the assisted mode turned off, the power used is similar (this was verified through two longer tests comparing assisted mode and non-assisted mode for 8 hours of transportation mode classification in non-scripted everyday operation). This is because the assisted service is used to improve the time to first fix (TTFF) for the GPS. In assisted mode, information about the currently seen cell towers is sent to an assisted server which then uses this data to provide signal parameters, which are a function of the mobile phone’s location, to reduce the GPS TTFF period. Based on observations with the Nokia n95, the assisted service is typically engaged when the transportation mode classifier is initially started (first time a fix of a GPS is required - often referred to as a cold start) and in some occasions if a GPS lock has not been obtained and the location of the phone has changed significantly or if cell tower information is lost and re-obtained. In general, these assisted server requests occur infrequently during operation, and even when the server is engaged the increased power usage (0.718 watts) occurs for twenty seconds on average (enough time to exchange information between the phone and the server).

5.7 Energy-Aware Detection

Our objective is to create a transportation mode classifier that captures the behavior of individuals when they are outside. Hence, it would be more energy efficient to only have the transportation mode classifier run when an individual is actually outdoors. The classifier can be turned off once GPS locks are lost for a period of time (indicating that the user is indoors). But the most effective method to determine when the user is outdoors again is by sampling the GPS and analyzing the presence and quality of location information. Unfortunately, uniformly sampling the GPS for this purpose is power hungry (0.407 watts). Thus, instead of uniformly sampling the GPS receiver once the user loses a lock for a period of time, we propose a triggered approach that relies on changes in the primary GSM cell tower.

The basic intuition of this method is that when individuals go from an indoor setting to an outdoor setting they typically move in terms of distance. Movement triggers the phone to lock onto different GSM cell towers, and attempting to sample the GPS when only changes occur to the primary GSM cell tower would be more efficient in terms of energy usage then blindly uniformly sampling the GPS. Essentially, changes in GSM cell towers are being used to determine the start of outdoor trips. Sampling and processing the GSM cell tower is a low energy cost operation with only 0.056 watts being used, which is 0.002 watts more than the idle state. In the GSM triggered based method, filtering is employed to eliminate the "ping pong" effect where the connected cell tower changes back and forth when stationary. Essentially, the filter clusters the pairs of cell tower identifiers that exhibit this behavior in a high frequency [Ravi et al. 2008].

To test the performance of the GSM triggered approach, day traces from sixteen individuals, who labeled indoor/outdoor status and collected GSM cell tower data every 1 second, was analyzed. The total time of the day trace data collection was on average 23.2 hours with a minimum time of 20.7 hours and a maximum time of 26.8 hours. The corresponding outdoor time is on average 3.09 hours with a maximum of 12.0 hours and a minimum of 0.93 hours. The number of outdoor events is 114. The GPS lock period is set to 60 seconds, and we assumed that locks would be successful when triggered outdoors during that time period.

The results of the GSM triggered approach show that, for the sixteen users, the average percentage of outdoor time identified is 91.5%, the average number of GPS checks is 68, and the total number of missed outdoor events is only 10. The missed events were on average less than 2 minutes with the longest being only 7 minutes. Most of the missed events were cases where the user stepped outdoors and then back indoors rapidly. In our application space, these events are not as important because they do not represent a significant transportation event. Comparing the GSM triggered approach to optimally uniformly sampling the GPS to achieve at least the same amount of outdoor time identified results in a 12.4% energy savings, even with the additional energy needed to obtain cell tower information, due to fewer GPS checks (73 on average) that needed to be made. The optimal uniform sampling rate is obtained by analyzing all data to figure out the ideal sampling rate based on the requirements for the amount of outdoor time needed to be represented, and since analyzing data a priori is not possible in a real application setting, the energy savings estimate is worst case (most conservative).

In summary, the GSM triggered approach is a very promising mechanism to save energy for obtaining transportation annotated mobility information. The individuals involved in our dataset were mainly in urban or sub-urban areas. In rural areas, where GSM cell towers are located with less density, larger portions of the start of the trip might not be recorded. However, the missing data can be approximated by analyzing where the individual lost GPS previously and filling in the gap accordingly. Overall, there exist other optimization techniques that could be employed to help make obtaining transportation mode annotated mobility information even more energy efficient. Our current method is just a start, and alternatives techniques should be further explored. For instance, accelerometer or WiFi changes along with GSM can be used for outdoor event triggers and the GPS sampling could be adapted using a back-off scheme [Nokia-Research 2009; Constandache et al. 2008] or by using contextual information such as time of day, day of week, and area of usage [Constandache et al. 2009]. We discuss such energy optimization techniques along with other future work in the discussion section.

6. DISCUSSION

The transportation mode classification results derived from the user base of sixteen individuals are very promising. The classification system is accurate regardless of position or orientation of sensors, and a generic classifier is feasible that works for different individuals without user-specific training or relying on external indexes. The results provide design strategies that industry could employ to include transportation mode classification as a first order service on mobile phones.

There are opportunities to tune the classification method even more. For instance, it would be interesting to investigate if user input, in terms of how the phone is used, could improve accuracy. One can imagine an initial survey being given to the user that asks types of transportation modes that are performed and then tuning the classifier based on this knowledge. Furthermore, location dependent user-specific models can be built to distinguish between the type of motorized transport (car, bus, train) used through in-situ experience sampling. This method could also help in determining when cases of underground transport have occurred, which is currently not supported by our system due to the lack of GPS information. Another area of further work is in making the classification system more robust to loss of sensor availability. If a GPS lock is not available or has poor quality, the classifier could backup to using WiFi or GSM features to help determine speed. Judging the validity of the WiFi and GSM features will be a key factor in how successful this method will be since the type of environment an individual is in affects their usefulness.

Also, research is warranted to make the classification method more energy efficient. Currently, this work focuses solely on efficiently detecting when an individual is outdoors to shutdown activity classification when indoors. But when a user is outside, it might not be necessary to sample the transportation mode classifier every second (this is, in fact, dependent on application requirements). For instance, the work of Krause et al. [2005] suggests that selective sampling techniques, such as ones based on entropy, can be employed while still achieving high accuracy. Alternatively, one might use low-power sensors to detect changes in transportation modes and then employ more accurate, higher power sensors to detect the exact type of mode change while still collecting location information in the background. This type of technique has been applied to reduce localization energy costs [Constandache et al. 2009] but not yet in the realm of transportation mode classification. Finally, the location module and classifier system could consider the cost of capturing and processing of sensing modalities to control the tradeoff between accuracy and energy consumption. For example, when the battery is low and the likelihood of having a recharge soon is small, it could be more helpful to obtain less confident location and transportation mode predictions for a longer period of time as opposed to using the high power sensors to get more accurate inferences for a shorter period [Banerjee et al. 2007; Ravi et al. 2008].

Another interesting area of work is how device differences (Nokia n95 vs Apple iPhone or TMobile G1) affects the makeup of the classifier. Our preliminary tests with alternative platforms indicate that the features and the classification setup is robust to such changes. In fact, the classifier created through the data collection obtained on the Nokia n95 was ported with high accuracy to the TMobile G1 platform by simply scaling the accelerometer data appropriately. But a more thorough characterization using GPS and accelerometers with different sensitivity characteristics and varied device integration positions needs to be completed.

7. CONCLUSION

We created a transportation mode classification system, employing a DT followed by a DHMM, that distinguishes between being stationary, walking, running, biking, and in motorized travel when an individual is outside using a mobile phone equipped with a GPS receiver and an accelerometer. The system is convenient for a user by not having strict position and orientation requirements while still achieving a high accuracy level of 93.6%, based on a dataset of one hundred and twenty hours of data from sixteen users. Furthermore, the classifier does not rely on external spatial indexes such as GIS data or historical usage patterns and works well even without user-specific training information.

REFERENCES

AGAPIE, E., CHEN, G., HOUSTON, D., HOWARD, E., KIM, J., MUN, M., MONDSCHEIN, A., REDDY, S., ROSARIO, R., RYDER, J., STEINER, A., BURKE, J., ESTRIN, D., HANSEN, M., and RAHIMI, M. 2008. Seeing our signals: Combining location traces and web-based models for personal discovery. In *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications*. ACM, New York, 6–10.

ANDERSON, I., AND MULLER, H. 2006. Practical activity recognition using GSM data. Technical Report CSTR-06-016, Department of Computer Science, University of Bristol.

ANTONSSON, E. AND MANN, R. 1985. The frequency content of gait. *J. Biomech. 18*, 1, 39–47.

BANERJEE, N., RAHMATI, A., CORNER, M., ROLLINS, S., AND ZHONG, L. 2007. Users and batteries: Interactions and adaptive energy management in mobile systems. In Lecture Notes in Computer Science vol. 4717. Springer-Verlag, Berlin, Germany.

BAO, L. AND INTILLE, S. 2004. Activity recognition from user-annotated acceleration data. In Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany, 1–17.

BODYMEDIA. 2008. Sensewear wms. http://www.sensewear.com/images/WMSbrochure.pdf.

BURKE, J., ESTRIN, D., HANSEN, M., PARKER, A., RAMANATHAN, N., REDDY, S., AND SRIVASTAVA, M. 2006. Participatory sensing. In *World Sensor Web Workshop*. ACM, New York, 1–5.

CARBONHERO. 2008. Carbon diem - Is that knowledge now. http://carbondiem.com.

CHEN, C., ANTON, S., AND HELAL, A. 2008. A brief survey of physical activity monitoring devices. University of Florida Tech Report MPCL-08-09.

CHEN, M., SOHN, T., CHMELEV, D., HAEHNEL, D., HIGHTOWER, J., HUGHES, J., LAMARCA, A., POTTER, F., SMITH, I., AND VARSHAVSKY, A. 2006. Practical metropolitan-scale positioning for gsm phones. In Lecture Notes in Computer Science 4206, Springer-Verlag, Berlin, Germany. vol. 225.

COMPUMINE. 2008. Evaluating a Classification Model. http://www.compumine.com/web/public/newsletter/20071/precision-recall.

CONSOLVO, S., KLASNJA, P., MCDONALD, D., AVRAHAMI, D., FROEHLICH, J., LEGRAND, L., LIBBY, R., MOSHER, K., AND LANDAY, J. 2008a. Flowers or a robot army?: Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing*. ACM New York, 54–63.

CONSOLVO, S., MCDONALD, D., TOSCOS, T., CHEN, M., FROEHLICH, J., HARRISON, B., KLASNJA, P., LAMARCA, A., LEGRAND, L., LIBBY, R., SMITH, I., and LANDAY, J. A. 2008b. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the 26th Annual Conference on Human Factors in Computing Systems*. ACM, New York, 1797–1806.

CONSTANDACHE, I., GAONKAR, S., SAYLER, M., CHOUDHURY, R., AND COX, L. 2009. EnLoc: Energy-efficient localization for mobile phones. IEEE Computer Society Press, Los Alamitos, CA. *Infocom Mini Conference*

CONSTANDACHE, I., SAYLER, M., GAONKAR, S., CHOUDHURY, R., AND COX, L. 2008. Energy-aware localization using mobile phones. *Mobisys Poster*. www.ee.duke.edu.

CROUTER, S., SCHNEIDER, P., KARABULUT, M., AND BASSETT JR, D. 2003. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med. Sci. Sports Exercise 35*, 8, 1455.

DENNING, T., ANDREW, A., CHAUDHRI, R., HARTUNG, C., LESTER, J., BORRIELLO, G., AND DUNCAN, G. 2009. BALANCE: Towards a usable pervasive wellness application with accurate activity inference. In *Proceedings of the 10th IEEE Workshop on Mobile Computing Systems and Applications (HotMobile 2009)*. IEEE Computer Society Press, Los Alamitos, CA.

DUDA, R., HART, P., AND STORK, D. 2000. *Pattern Clasification*. Wiley, New York.

ESPOSITO, F., MALERBA, D., SEMERARO, G., AND KAY, J. 1997. A comparative analysis of methods for pruning decision trees. In *Pattern Analysis and Machine Intelligence, IEEE Trans*. *19*. 476–491.

FÁBIÁN, Á., GYŐRBÍRÓ, N., AND HOMÁNYI, G. 2008. Activity recognition system for mobile phones using the MotionBand device. In *Proceedings of the 1st International Conference on Mobile Wireless Middleware, Operating Systems, and Applications*. ICST, Brussels, Belgium.

FARRINGDON, J., MOORE, A., TILBURY, N., CHURCH, J., AND BIEMOND, P. 1999. Wearable sensor badge & sensor jacket for context awareness. In *Proceedings of the 3rd International Symposium on Wearable Computers*. 107–113.

FROEHLICH, J., DILLAHUNT, T., KLASNJA, P., MANKOFF, J., CONSOLVO, S., HARRISON, B., AND LANDAY, J. 2009. UbiGreen: Investigating a mobile tool for tracking and supporting green transportation habits. *Computer Human Interaction (CHI)*. ACM, New York.

GANTI, R., JAYACHANDRAN, P., ABDELZAHER, T., AND STANKOVIC, J. 2006. SATIRE: A software architecture for smart AtTIRE. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*. ACM New York, 110–123.

GAONKAR, S., LI, J., CHOUDHURY, R., AND COX, L. 2008. Micro-blog: Sharing and querying content through mobile phones and social participation. In *Proceedings of the 6th International Conference on Mobile Systems, Applications and Services*. ACM, New York, 174–186.

GHISELLI, E. 1964. *Theory of Psychological Measurement*. McGraw-Hill, New York.

GHMM. 2008. The General Hidden Markov Model Library. http://ghmm.org.

GOERTZEL, G. 1958. An algorithm for the evaluation of finite trigonometric series. *Amer. Math. Monthly 65*, 1, 34–35.

HUYNH, T. AND SCHIELE, B. 2005. Analyzing features for activity recognition. In *Proceedings Conference on Smart Objects and Ambient Intelligence*. ACM, New York, 159–163.

IMPACTSPORTS. 2008. ePulse - Armband Heart Rate Monitor. http://impactsports.com.

JAAKKOLA, T., AND HAUSSLER, D. 1999. Exploiting generative models in discriminative classifiers. *Adv. Neural Info. Proc. Syst.* 487–493.

KAO, J., LAM, R., PONG, G., TALUKDAR, T., AND WONG, J. 2008. Ecorio - Track Your Mobile Carbon Footprint. http://ecorio.org.

KARABULUT, M., CROUTER, S., AND BASSETT, D. 2005. Comparison of two waist-mounted and two ankle-mounted electronic pedometers. *Europ. J. Applied Physiology 95*, 4, 335–343.

KERN, N., SCHIELE, B., AND SCHMIDT, A. 2003. Multi-sensor activity context detection for wearable computing. In Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 220–234.

KRAUSE, A., IHMIG, M., RANKIN, E., LEONG, D., GUPTA, S., SIEWIOREK, D., SMAILAGIC, A., DEISHER, M., AND SENGUPTA, U. 2005. Trading off prediction accuracy and power consumption for context-aware wearable computing. In *Proceedings of the 9th IEEE International Symposium on Wearable Computers*. IEEE Computer Society Press, Los Alamitos, CA. *20*, 25.

LAMARCA, A., CHAWATHE, Y., CONSOLVO, S., HIGHTOWER, J., SMITH, I., SCOTT, J., SOHN, T., HOWARD, J., HUGHES, J., POTTER, F., TABERT, J., POWLEDGE, P., BORRIELLO, G., and SCHILIB, B. 2005. Place lab: Device positioning using radio beacons in the wild. In *Proceedings of PERVASIVE*. Vol. 3468. Springer-Verlag, Berlin, Germany, 7.

LESTER, J., CHOUDHURY, T., AND BORRIELLO, G. 2006. A practical approach to recognizing physical activities. In Lecture Notes in Computer Science, vol. 3968, Springer-Verlag, Berlin, Germany, 1.

LESTER, J., CHOUDHURY, T., KERN, N., BORRIELLO, G., AND HANNAFORD, B. 2005. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2005)*.

LESTER, J., HURVITZ, P., CHAUDHRI, R., HARTUNG, C., AND BORRIELLO, G. 2008. MobileSense - Sensing modes of transportation in studies of the built environment. In *Proceedings of the International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems*. ACM, New York.

LI, K., SOHN, T., HUANG, S., AND GRISWOLD, W. 2008. Peopletones: A system for the detection and notification of buddy proximity on mobile phones. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*. ACM, New York.

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

MARMASSE, N., SCHMANDT, C., AND SPECTRE, D. 2004. WatchMe: communication and awareness between members of a closely-knit group. In Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany, 214–231.

MARTINEZ, A. AND KAK, A. 2001. PCA versus LDA. *Patt. Anal. Mach. Intell. IEEE Trans. 23*, 2, 228–233.

MILUZZO, E., LANE, N., EISENMAN, S., AND CAMPBELL, A. 2007. Cenceme-injecting sensing presence into social networking applications. In Lecture Notes in Computer Science. vol. 4793. Springer-Verlag, Berlin, Germany, 1.

MILUZZO, E., LANE, N., FODOR, K., PETERSON, R., LU, H., MUSOLESI, M., EISENMAN, S., ZHENG, X., AND CAMPBELL, A. 2008. Sensing meets mobile social networks: The design implementation and evaluation of the cenceMe application. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*. ACM, New York. 337–350.

MOHAN, P., PADMANABHAN, V., AND RAMJEE, R. 2008. Trafficsense: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the Conference on Embedded Networked Sensor Systems (SenSys)*. ACM, New York, 323–336.

MUN, M., ESTRIN, D., BURKE, J., AND HANSEN, M. 2008. Parsimonious mobility classification using GSM and WiFi traces. In *Proceedings of the 5th Workshop on Embedded Networked Sensors (HotEmNets)*.

MUN, M., REDDY, S., SAMANTA, V., SHILTON, K., HOWARD, E., LE, D., RAMANTHAN, N., BURKE, J., ESTRIN, D., HANSEN, M., AND SRIVASTAVA, M. 2009a. CycleSense: Sensing and mapping for better biking. In *Proceedings of the Active Living Conference*.

MUN, M., REDDY, S., SHILTON, K., YAU, N., BODA, P., BURKE, J., ESTRIN, D., HANSEN, M., HOWARD, E., AND WEST, R. 2009b. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*. ACM, New York, 55–68.

NIKE. 2008. Apple Computer, Inc., Nike+ iPod Sport Kit. http://apple.com/nike.

NOKIA. 2008. Nokia n95. http://www.nseries.com.

NOKIA-RESEARCH. 2009. Nokia simple context service. https://alpha.nokoscope.com/eb2/.

OMRON. 2008a. Omron hj-150. http://www.omronhealthcare.com/product/detail.asp?p=1129&t=187.

OMRON. 2008b. Omron hj-720itc. http://www.omronhealthcare.com/product/1132-187-fitness-diagnostics-pocket-pedometer-with-pc-software-hj-720itc.

PATTERSON, D., LIAO, L., FOX, D., AND KAUTZ, H. 2003. Inferring high-level behavior from low-level sensors. In Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 73–89.

PETERSON, N., ESTRIN, D., AND BURKE, J. 2009. Project footstep - Linking physical activity with location. http://urban.cens.ucla.edu/projects/footstep.

PHAM, N., AND ABDELZAHER, T. 2008. Robust dynamic human activity recognition based on relative energy allocation. In Lecture Notes in Computer Science, vol. 5067, Springer-Verlag, Berlin, Germany, 525–530.

RANDELL, C., AND MULLER, H. 2000. Context awareness by analysing accelerometer data. In *Proceedings of the 4th International Symposium on Wearable Computers*. IEEE Computer Society Press, Los Alamitos, CA, 175–176.

RAVI, N., SCOTT, J., HAN, L., AND IFTODE, L. 2008. Context-aware battery management for mobile phones. In *Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom2008)*. IEEE Computer Society Press, Los Alamitos, CA, 224–233.

REDDY, S., SAMANTA, V., BURKE, J., ESTRIN, D., AND SRIVASTAVA, M. 2009a. Walkability - Safe Walkable Streets in Modern Cities. http://urban.cens.ucla.edu/projects/walkability.

REDDY, S., SHILTON, K., BURKE, J., ESTRIN, D., HANSEN, M., AND SRIVASTAVA, M. 2009b. Using context annotated mobility profiles to recruit data collectors in participatory sensing. In *Proceedings of the Symposium on Location and Context Awareness (LOCA)*. Springer-Verlag, Berlin, Germany.

SAPONAS, T., LESTER, J., FROEHLICH, J., FOGARTY, J., AND LANDAY, J. 2008. iLearn on the iPhone: Real-time human activity classification on commodity mobile phones. University of Washington CSE Tech Report UW-CSE-08-04-02.

SCHNEIDER, P., CROUTER, S., AND BASSETT, D. 2004. Pedometer measures of free-living physical activity: comparison of 13 models. *Med. Sci. Sports Exerc 36*, 2, 331–335.

SKYHOOK. 2009. Skyhook Wireless: Coverage. http://www.skyhookwireless.com/howitworks/coverage.php.

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. In Lecture Notes in Computer Science, vol. 4206, Springer-Verlag, Berlin, Germany, 212.

SUN, M., AND HILL, J. 1993. A method for measuring mechanical work and work efficiency during human activities. *J. Biomech. 26*, 3, 229–241.

TAPIA, E., INTILLE, S., AND LARSON, K. 2004. Activity recognition in the home using simple and sbiquitous sensors. In Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany, 158–175.

WELBOURNE, E., LESTER, J., LAMARCA, A., AND BORRIELLO, G. 2005. Mobile context inference using low-cost sensors. In *Proceedings of the 1st International Symposium on Location and Context Awareness*. Springer-Verlag, Berlin, Germany, 254–263.

WIGLE. 2009. WiGLE - Wireless Geographic Logging Engine. http://wigle.net/gps/gps/Map/ onlinemap2/.

WINTER, D. 1990. *Biomechanics and Motor Control of Human Movement*. Wiley, New York.

WITTEN, I. AND FRANK, E. 2005. *Data mining: Practical machine learning tools and techniques*. Morgan-Kaufmann, San Francisco, CA.

YU, L. AND LIU, H. 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th International Conference on Machine Learning*. 856–864.

ZHENG, Y., LI, Q., CHEN, Y., XIE, X., AND MA, W. 2008. Understanding mobility based on GPS data. In *Ubiquitous Computing*. ACM New York, 312–321.

Received December 2008; revised April 2009; accepted June 2009