Performing Activity Recognition Using an Android Smartphone

Stefan Dernbach
Whitworth University
Department of Computer Science
300 W Hawthrone
Spokane, WA 99352

Barnan Das, Brian Thomas, Diane Cook
Washington State University
School of Electrical Engineering and Computer Science
Pullman, WA 99164

***Abstract*-This paper investigates the ability to perform activity recognition on simple (single action) and complex (multi-action) activities using the accelerometer and orientation data collected from a mobile phone. Data collection was performed using an Android smartphone for 8 men and 2 women. Features extracted from the data were used to train and test a sample of machine learning algorithms. The performance of the algorithms was measured over various window lengths of time.**

I. Introduction

Human activity recognition is an important area of machine learning research because of its real-world uses. Automated activity recognition reduces the necessity for humans to oversee difficulties individuals might have performing activities, such as falling when they try to get out of bed. Activity recognition can also be used in conjunction with pattern recognition to determine changes in a subject's routine. For these reasons the technology has many potential uses in healthcare and eldercare.

Two methods of collecting data for performing activity recognition have been extensively researched. The first method relies upon environmental sensors to track features such as motion, location, and object interaction. Alternatively, the second method uses a network of sensors attached to the human body to track the acceleration of specific limbs as well as the body as a whole. Both of these methods have demonstrated impressive results.

A major hurdle in implementing these systems outside of trials is how unnatural the sensors are. Environmental sensors are generally bulky and costly. They also must be wired or have their batteries maintained. Both cases involve a large investment into setting up and maintaining the system. Body sensors require daily effort from the user to wear and maintain them or else they are useless for collecting data. Additionally they are bulky and require batteries, two factors that reduce the likelihood of constant use by the user.

This paper describes the effectiveness of using the accelerometer and gyroscope of a smartphone as a more natural alternative to a series of body sensors. While many older adults are currently not likely to carry smartphones, younger people are increasingly likely to carry a mobile phone on them. These people represent the next generations of elders. So while a phone may still be an unnatural accessory for older people it is becoming increasingly less so. Using a phone as the primary device for data collection increases the likelihood of data coverage and represents a minimal cost and maintenance commitment to the user.

Data for the training and testing recognition algorithms was collected from ten subjects who were instructed to wear the phone naturally as they performed various activities.

II. Background

Previous activity recognition work has focused on using a network of environmental or body sensors. While activities and restrictions have varied between studies, most have provided promising results.

A. Environmental Sensors

Tapia, et al. [1] collected data in two homes containing 77 and 84 separate state-change sensors, respectively. Seventeen separate activities were monitored and activities could overlap. The goal was to determine when an activity began, what activity it was, and how long it lasted. Data was self-annotated by subjects in the study using the context-aware experience sampling tool [2]. The tool is a PDA that subjects carried with them that would prompt them to respond with what activity they are performing. The study found a large range (from 25% to 89%) in the accuracy of detecting each activity. Overall, a peek accuracy of 38.79% was reported for correctly detecting an activity using a multiclass naïve Bayes classifier.

Another study let by Philipose, et al. [3] used radio-frequency-identification (RFID) tags which attach to objects and trigger when near an RFID reader worn on the subject's hand. The study involved fourteen activities including oral hygiene, using appliances, and infant care. Activities were correctly recognized as occurring 88% of the time and the correct activity was identified 73% of the time.

Yet another study employed environmental sensors to perform long-term activity recognition in participant homes as they performed their normal daily routines [4]. This study also demonstrated that activity models could be transferred across multiple spaces to recognize activities with no training data in a new setting.

B. Body Sensors

Other researchers have focused on using wearable sensors to perform activity recognition. Bao and Intille [5] used a series of five biaxial accelerometers sampling at a rate of 76.25 Hz. Sensors were placed on the left bicep, right wrist, left quadriceps, right ankle, and right hip. The accelerometers couldn't be wired together without restricting the subject's movement so each kept its own time. Subjects performed twenty different activities ranging from walking to folding laundry to strength training. Data on these activities was collected while subjects completed an obstacle course. Subjects also performed activities in a controlled setting. For feature extraction, a window size of 512 samples was used (6.7 seconds at 76.25 HZ). An accuracy of 84% was reported using decision tree classifiers, while the accuracy of individual activities ranged from 41.42% (stretching) to 97.49% (working on computer).

A second experiment conducted by Ravi, et al. [6] used a single accelerometer mounted onto the pelvic region of subjects to collect data on eight activities: Standing, walking, running, climbing up stairs, climbing down stairs, sit-ups, vacuuming, and brushing teeth. The accelerometer transmitted data over Bluetooth to a PDA at a rate of 50 Hz. The data was annotated by the researcher as it occurred. Four features: mean, standard deviation, energy, and correlation, were extracted from the raw accelerometer data in window lengths of 256 samples (roughly 5.1 seconds at 50 Hz). Using cross validation, peek accuracy was 99.57% for recognizing a single subjects activities and 99.82% for multiple subjects. This accuracy was achieved using plurality voting which selects the most common prediction from five classifiers: decision tables, decision trees, K-nearest neighbors, SVM, and naïve Bayes.

C. Discussion

Previous activity recognition experiments have had a large range in reported accuracies. Some of the differences between experiments include the number of activities and the complexity of the activities. A larger number of activities means more overlap in activity features and makes it harder to discern exactly which activity is being performed. This difficulty is compounded in environmental sensor studies when it first has to be determined if any activity is taking place at all. Activities can also be concurrent, performed at the same time, or interleaved where one activity is broke up by a other concurrent activities [7].

Simple activities, such as walking, can be represented as a single repeated action: taking a step forward; whereas complex activities, such as taking medication, involve multiple actions: opening a cupboard, taking out pills, swallowing, and returning the remaining pills. Other complex activities may involve simultaneous or overlapping actions. These traits decrease the ability to reduce these actions down to discrete features.

Another factor that affects the performance of activity recognition systems is how general the system as a whole is. Does it classify all driving as a single category or is highway driving separate from city driving? Also of note is how many subjects and locations are involved with the system. The algorithm can be tailored to a specific individual in a specific environment or be fitted for multiple subjects in any number of environments.

III. DESIGN

The activity recognition is performed as a three-part process. First data is collected and pooled from multiple participants, than features are extracted from that data, and finally machine learning algorithms are trained and tested.

Subjects wore an Android smartphone that contained a triaxial accelerometer and gyroscope. The location and orientation of the phone was not standardized. Subjects input the task they were about to perform on the phone and then handled starting and stopping the data collection (Fig.1). Subjects also included the location they were wearing the phone purely as a reference.

A. Activities

Activities were divided into two categories: simple and complex. Simple activities consist of a single repeated action whereas complex activities are the compilation of a series of multiple actions. Subjects performed simple activities in variable amounts and in various environments; the action, location, and length of performance was not controlled. The simple activities consisted of:

- Biking
- Climbing
- Driving
- Lying

Fig. 1. The application used for collecting activity data.

- Phone Not on Person
- Running
- Sitting
- Standing
- Walking

Every subject performed each complex activity four times in the same apartment. The complex activities had defined starting and finishing points and lasted until the subject completed them. The complex activities consisted of:

- Cleaning: Each subject wiped down the kitchen counter and sink.
- Cooking: Each subject simulated cooking by heating a bowl of water in the microwave and pouring a glass of water from a pitcher in the fridge.
- Medication: Each subject retrieved pills from the cupboard and sorted out a week's worth of doses.
- Sweeping: Each subject swept the kitchen area.
- Washing hands: Each subject washed their hands using the soap at the kitchen sink.
- Watering Plants: Each subject filled a watering can and water three plants in two rooms.

B. *Feature Extraction*

Raw data is collected as a series of instances containing a timestamp, three values corresponding to acceleration along the x-axis, y-axis, and z-axis, and a second set of three orientation values representing azimuth, pitch, and roll. Rather than a set sampling rate, the accelerometer in an Android phone triggers an event whenever the accelerometer values change. The rate of events can be set to one of four thresholds: fastest, game, normal, and UI, with fastest being the fastest sampling rate and UI being the slowest. The phone used for this experiment was set to fastest. The sampling rate varies because of this but can reach a maximum of 80 Hz. The three axes of acceleration are dependent upon the orientation of the phone. The x-axis runs parallel to the width of the phone, the y-axis runs the length of the phone, and the z-axis runs perpendicular to the face of the phone, as shown in Fig. 2.

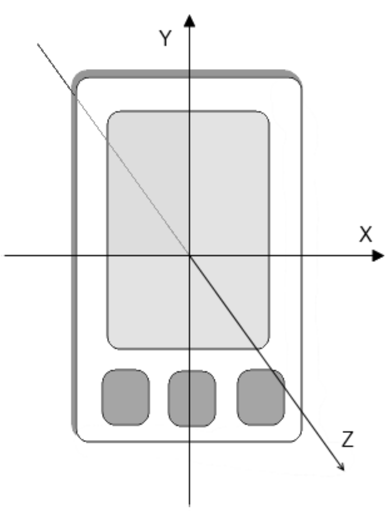

Fig. 2. Axes of Acceleration Relative to Phone

Raw data is processed to normalize the acceleration axes so that the x-axis, y-axis, and z-axis run north and south, east and west, and up and down respectively. The orientation data is left as is.

Features (shown in Table I) are then extracted from a sliding window of varying lengths of time. Performance was tested on window lengths of one, two, four, eight, twelve, and sixteen seconds. Windows always overlapped by one half of the window length, e.g., a four second window slides over two seconds at a time. Thus, each window is a single instance, but any given data point contributes to two instances. This method has been shown to be effective in earlier work using accelerometer data [5, 6].

C. *Classification*

The WEKA machine learning toolkit [8] was used to test classifiers using the features extracted from the raw data set. Six different classifiers were tested: multi-layer perceptron, naïve Bayes, Bayes network, decision table, best-first tree, and K-star. The accuracy of the classifiers was tested using cross-validation with ten folds. WEKA was also used remove orientation features when testing classifiers using only acceleration values.

IV. RESULTS

Classification was tested for the selected classifiers using six different window lengths: one, two, four, eight, twelve, and sixteen seconds as well as three activity sets: simple, complex, and combined. Overall, accuracy for simple tasks was high across the board for both various classifiers (Fig. 3) as well as across a range of window lengths (Fig. 4) with only the naïve Bayes classifier performing under 90% accuracy.

This trend did not continue when complex activities were added to the set. Shorter window frames performed significantly better than longer ones. This result is surprising as a shorter window is less likely to account for more than a single one of the actions involved in a complex

TABLE I
Feature List

Feature	Accelerometer	Orientation
Mean	X, Y, Z	Azimuth, Pitch, Roll
Min	X, Y, Z	Azimuth, Pitch, Roll
Max	X, Y, Z	Azimuth, Pitch, Roll
Standard Deviation	X, Y, Z	Azimuth, Pitch, Roll
Zero-Cross	X, Y, Z	
Correlation	X/Y, X/Z, Y/Z	

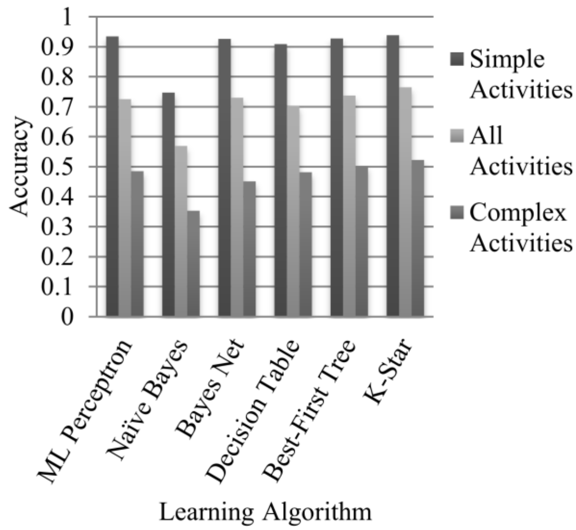

Fig. 3. Accuracy of Various Classifiers

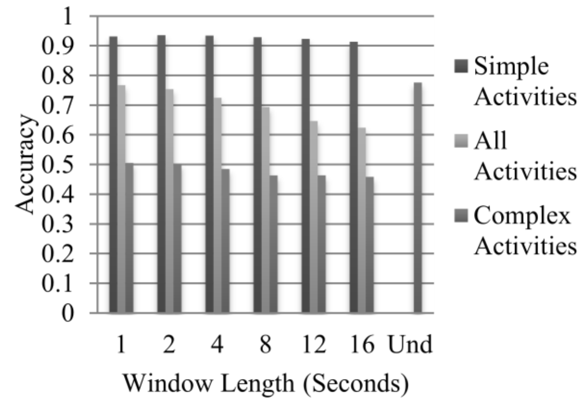

Fig. 4. Accuracy of Window Lengths

activity. The set of just complex activities barely achieved 50% accuracy; blind guessing achieves 17% accuracy. Table II shows the confusion matrix for the combined activity set. The matrix shows how heavily the misclassifications lean towards complex activities. This is not surprising alone but does show the resilience of simple tasks against being misclassified when more activities are added.

When looking only at complex activities in Table III, cooking, medication, and sweeping were all (correctly and incorrectly) classified far more often than cleaning, washing hands, or watering plants. The former activities all involved little movement in the experiment whiles the latter (with the exception of washing hands) contained much more movement from one area to another.

Overall, The K-Star classifier was the most accurate classifier performing best for all three activity sets. The naïve Bayes classifier performed worst in all three cases.

The complex activity set was also tested without using a sliding window for feature extraction. Instead, an entire activity acted as a single instance. This greatly improved performance from 52% to 78% accuracy. It is worth noting that the amount of data used to train and test the classifier was drastically reduced when the data is treated this way to forty instances per activity.

Unfortunately, there are several issues with performing recognition in this method. This process was possible because the subjects input the start and stop times of the activities they performed. A fully automated recognition system would need some way of determining the start and end of an activity rather than relying on the user. Additionally, unlike a sliding window which can do recognition in pseudo-real-time, this method requires an activity to be completed before it can be recognized. The results do however; represent the potential for improving the recognition system.

Fig. 5 shows the results of classifying each activity set with and without features extracted from orientation. Previous studies have generally relied solely on acceleration data and did not take the orientation of the sensors into account after beginning an activity. In this case, orientation data represented an average of a 10-12% increase in accuracy over pure acceleration data. Given the format of

TABLE II
Confusion Matrix for Combined Activity Set
(Multi-layer Perceptron, 1-Second Window)

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	← classified as
1056	52	29	13	20	20	9	5	21	34	66	26	169	33	50	a = Biking
28	694	18	18	12	33	21	33	31	11	7	12	28	8	13	b = Climbing
3	25	3228	2	3	5	1	42	11	0	12	1	2	1	5	c = Driving
27	34	10	1413	9	2	2	1	5	2	4	16	2	2	1	d = Lying
0	3	1	1	3360	0	1	1	0	0	0	0	1	0	0	e = Not on Person
1	16	3	4	1	780	0	1	1	4	6	3	16	7	5	f = Running
27	51	29	6	1	6	1107	5	7	8	1	4	4	10	2	g = Sitting
1	6	0	5	3	6	0	748	0	0	1	1	1	0	2	h = Standing
9	68	10	7	2	8	17	1	1472	8	13	14	40	2	23	i = Walking
0	0	1	0	0	1	0	0	8	190	243	108	209	2	11	j = Cleaning Kitchen
1	1	0	5	0	4	1	0	5	123	774	160	448	18	123	k = Cooking
0	5	0	1	0	3	0	1	19	77	125	1103	381	27	53	l = Medication
0	2	0	2	0	6	0	0	11	116	348	176	1257	21	68	m = Sweeping
0	1	1	0	1	2	0	0	4	42	96	62	149	56	38	n = Washing Hands
0	0	0	1	1	8	2	0	8	30	125	70	171	24	245	o = Watering Plants

TABLE III
Detailed Accuracy by Class for Complex Activities Set
(Multi-layer Perceptron, 1-Second Window)

Tp Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.182	0.030	0.418	0.182	0.254	0.737	Cleaning Kitchen
0.536	0.189	0.453	0.536	0.491	0.766	Cooking
0.673	0.136	0.614	0.673	0.642	0.828	Medication
0.600	0.235	0.489	0.600	0.539	0.761	Sweeping
0.153	0.016	0.390	0.153	0.219	0.693	Washing Hands
0.318	0.036	0.474	0.318	0.381	0.735	Watering Plants
0.506	0.147	0.496	0.506	0.489	0.769	Weighted Average

this experiment it is likely that this is in part due to the fact that the starting position and orientation of the phone was not standardized. Features extracted from orientation helped overcome this deficiency.

V. Conclusions

Simple activities can be recognized with very high accuracy using only a single smartphone carried naturally. Performance was over 93% using a multi-layer perceptron network and a two second time window. The length of the window had very little effect on results for simple activities which implies that it can be reduced for recognizing short activities or extended as needed. Activity sets that included complex activities did not perform as well but still achieved over 50% accuracy. Simple activities retained their high classification accuracy even when paired with complex activities.

The results for simple activities are on par with previous work on body sensors [4]. This shows a lot of promise for using mobile phones as an alternative to dedicated accelerometers. The recognition of complex activities was also similar to that of the less recognizable activities in Ref. [3]. While 50% accuracy is not high enough for many real-world uses of activity recognition, it does show that a phone could be effective as part of a data collection system for recognizing complex activities even if it cannot function as a standalone system.

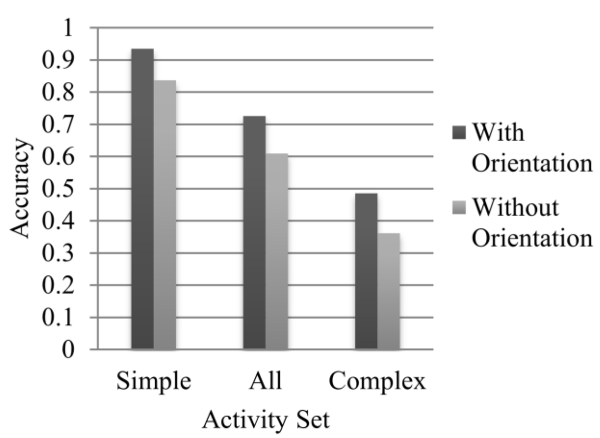

Fig. 5. Accuracy of Various Classifiers

VI. Future Work

An area of further research is to determine the effectiveness of classification using a lower sampling rate with the goal of reducing the strain on the phones battery. Evaluation should also be performed on subjects who did not contribute the recognition model as well as determining the effectiveness of a model tuned to only a single subject.

There remains plenty of work to do to improve the accuracy of activity recognition. One approach that merits further research is the combination of a mobile phone with environmental sensors. The combination of the two sensor types provides for detailed data on the subjects movement, location, and interactions. Data from object sensors may be essential in categorizing complex activities such as washing hands or watering plants. For example, watering plants could be seen as a combination of walking, standing, interacting with objects such as the sink and a watering can, and being in a particular area.

Additionally, recognizing a complex activity as a series of simple activities holds promise. Constructing a vector of a sequence of simple activities using the same machine learning techniques described in this paper and then using that vector to learn a complex activity may provide a higher accuracy of recognition than has previously been achieved.

Aknowledgments

This work was supported by the National Science Foundation's REU program under grant number IIS-0647705

References

[1] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson, "Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors," *Proceedings of PERVASIVE*, pp. 158-175, 2004.

[2] Stephen S. Intille, John Rondoni, Charles Kukla, Isabel Ancona, and Ling Bao, "A Context-aware Experience Sampling Tool," *Proceedings of the Conference on Human Factors and computing Systems: Extended Abstracts*, ACM Press, 2004.

[3] Philipose M, Fishkin K, Perkowitz M, Patterson D, Fox D, Kautz H, Hahnel D, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, vol. 3, pp. 50-57, 2004.

[4] D. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intelligent Systems*, to appear.

[5] Ling Bao and Stephen S. Intille, “Activity Recognition from User-Annotated Acceleration Data,” *Pervasive Computing*, LNCS 3001, pp. 1-17, 2004.

[6] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Machael L. Littman, “Activity Recognition from Accelerometer Data,” *Proceeding of the National Conference on Artificial Intelligence*, vol. 20, PART 3, pp. 1541-1546, 2005.

[7] Tao Gu, Zhanqing Wu, Xianping Tao, Hung Keng Pung, and Jian Lu, “epSICAR: An Emerging Patterns based Approach to Sequential, Interleaved and Concurrent Activity Recognition,” *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, 2009.

[8] I Witten and E. Frank, “Data Mining: Practicial Machine Learning Tools and Techniques with Java Implementations,” Morgan Kaufmann, 1999.