

Simple and Complex Activity Recognition Through Smart Phones

Stefan Dernbach
Department of Computer Science
Whitworth University
Spokane, USA
sdernbach12@my.whitworth.edu

Barnan Das, Narayanan C. Krishnan,
Brian L. Thomas, Diane J. Cook
School of EECS
Washington State University
Pullman, USA
{barnandas, ckn, bthomas1, djcook}@wsu.edu

Abstract—Due to an increased popularity of assistive healthcare technologies activity recognition has become one of the most widely studied problems in technology-driven assistive healthcare domain. Current approaches for smart-phone based activity recognition focus only on simple activities such as locomotion. In this paper, in addition to recognizing simple activities, we investigate the ability to recognize complex activities, such as cooking, cleaning, etc. through a smart phone. Features extracted from the raw inertial sensor data of the smart phone corresponding to the user’s activities, are used to train and test supervised machine learning algorithms. The results from the experiments conducted on ten participants indicate that, in isolation, while simple activities can be easily recognized, the performance of the prediction models on complex activities is poor. However, the prediction model is robust enough to recognize simple activities even in the presence of complex activities.

Keywords-activity recognition; accelerometer; smart environments; smart phone

I. INTRODUCTION

Human activity recognition is an important area of machine learning research because of its real-world applications. Automated activity recognition reduces the necessity for humans to oversee difficulties individuals (especially older adults) 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 in constrained laboratory settings.

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 smart phone as a more natural alternative to a combination of body sensors. While many older adults are currently not likely to carry smart phones, 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.

Current generation smart phones are equipped with a variety of sensors such as GPS sensors, microphones, image sensors (camera), light sensors, proximity sensors, inertial sensors (accelerometers and gyroscopes), and direction sensors (compass). The small form factor of the smart phones coupled with its ubiquity and the substantial computing power makes them an effective tool for understanding the current state of the smart phone user. In this paper, we explore the use of inertial sensors (accelerometers and gyroscopes) in the smart phone to identify the activity that a user is performing by mining the sensor data.

We have chosen an Android smart phone as the platform for the project for multiple reasons – the Android operating system is open-source and easily programmable and more importantly, its dominance in the smart phone market. A Samsung Captivate™ smart phone is used as the device running Android 2.1.

Activity recognition from inertial sensor data is not a new problem. There have been many approaches proposed in the literature as discussed in Section II. Our approach is unique from the existing methods in many ways. First and foremost, no additional body sensor is used on the subject. Secondly, the

subject's body location where the phone should be placed and the orientation of the phone are not predetermined. In addition, we perform activity recognition on complex activities such as, cooking, cleaning, etc. Unlike other works, where environmental sensors are used to recognize complex activities, we make an endeavor to use accelerometer and gyroscope data to achieve the same goal. To our knowledge, there has been no previous work in this direction.

We approach the activity recognition task as a supervised machine learning problem. The data collection process, the chosen set of activities and other details of the design of the experiment are presented in Section III. The experiments and the results obtained are discussed in Section IV. Section V concludes the paper and a discussion on future directions is presented in Section VI.

II. RELATED WORK

Advances in ubiquitous and pervasive computing have resulted in the development of a number of sensing technologies for capturing information related to human physical activities. The different approaches to activity recognition can be categorized based on the underlying sensing mechanism – network of environmental sensors of body area networks.

Environmental sensor-based activity recognition has received significant focus in the recent years. This is a promising approach for recognizing activities that are not easily distinguishable by body movement alone. Motion sensors, door contact sensors, object sensors RFID tags and video cameras are some of the most commonly used environmental sensors for gathering activity related information [1].

Wearable accelerometers have proved to be effective sensors for human activity recognition. Some of the earliest work on wearable sensor based activity recognition used multiple accelerometers placed on different parts of the body. Bao and Intille [4] used a series of five biaxial accelerometers placed on the left bicep, right wrist, left quadriceps, right ankle, and right hip for recognizing twenty different activities ranging from walking to folding laundry to strength training. Ravi et. al. [5] use 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 peak 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. Tapia et. al. [7] collected data from five accelerometers placed on various body locations for implementing a real-time system to recognize thirty gymnasium activities. Krishnan et. al. [8] collected data using two accelerometers for recognizing locomotion activities in real-time using adaptive boosting on decision stumps.

Other research has explored the use of multiple kinds of on-body sensors for activity recognition. Maurer et. al. [10] use accelerometers, temperature sensors and microphones for recognizing human locomotion. Their research also analyzed multiple time domain feature sets for activity recognition. Lee and Mase [11] proposed a system for recognizing activities

using information about the user's location and inertial sensors such as accelerometers and gyroscopes. Subramanya et. al. [12] also proposed a similar approach involving accelerometers, microphones, light sensors, thermometers, barometric pressure sensors and GPS sensors.

There have been a few studies similar to the one proposed in the paper that use commercially available mobile devices to collect data for activity recognition. Kwapisz et. al. [13] use an Android-based smart phone for recognizing very simple activities such as walk, jog, climb up and down the stairs, sit and stand. Yang [14] developed an activity recognition system using the Nokia N95 cell phone for distinguishing between different locomotion. Brezmes et. al. [15] proposes a subject dependent real time activity recognition system again using the Nokia N95 smart phone. Hache et. al. [16] use an accelerometer integrated with a blackberry Bold 9000 platform for detecting changes in the state of the subject caused by starting/stopping and postural changes in activities. Khan et. al. [17] use kernel discriminant analysis for recognizing very simple activities such as walking, up and down the stairs, running and resting on data collected from Samsung Omnia. Zhang et. al. [18] use an HTC smart phone for recognizing again simple activities using a support vector machines.

Mobile devices offer a number of advantages including unobtrusiveness of the system and not requiring any additional equipment for data collection or computing that make them an attractive platform for activity recognition. We build on these approaches and extend them to test the ability of mobile sensing and computing platforms for recognizing simple and complex activities such as watering plants and sweeping.

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.

III. DESIGN

In this section we describe the design of our experiments for performing activity recognition from smart phone. We begin this with a description of the data collection process and the set of activities considered for this work. We then describe the features that were extracted from the sensor data and finally discuss the machine learning algorithms that were used for performing the recognition task.

A. Data Collection

The data collection was done by performing experiments on ten undergraduate students who participated in National Science Foundation's REU program at Washington State University. Subjects wore an Android 2.1 operating system-based Samsung Captivate™ smart phone that contained a tri-axial accelerometer and gyroscope. The location and orientation of the phone was not standardized and was left to the convenience of the subject. However, orientation information was taken into consideration while making comparative study in Section IV. Each subject carried the

phone while performing the different activities. We created an application that stored the sensor data while the subject performed the activities. The application permitted us to control the sensor type from which the data was collected along with the sampling rate of the sensor.

Subjects controlled the data that was being collected through the application (Figure 1) we created that executed on the phone. The application allows the subjects to input the activity that they are about to perform along with the ability to start and stop recording the sensor data. The application also allows the subjects to input the location they were wearing the phone. Though, we collected this information, we did not use it the current study.



Figure 1: Application used for activity data collection

B. 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 included *biking*, *climbing stairs*, *driving*, *lying*, *running*, *sitting*, *standing* and *walking*. In addition, we also wanted to test the possibility of detecting the smart phone not being worn by the subject. We call this activity as *phone not on person*. While some of the simple activities also feature on the list of activities of other researchers, our list contains additional activities such as *driving*, that not has been studied before.

Every subject also performed a set of complex activities wearing the smart phone. The subjects repeated execution of these complex activities four times. The complex activities had definite starting and finishing points and lasted until the subject completed them. The complex activities consisted of:

- *Cleaning*: Subject wiped down the kitchen counter top and sink.

- *Cooking*: 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*: Subject retrieved pills from the cupboard and sorted out a week's worth of doses.
- *Sweeping*: Subject swept the kitchen area.
- *Washing Hands*: Subject washed hands using the soap at the kitchen sink.
- *Watering Plants*: Subject filled a watering can and watered three plants in two rooms.

C. 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 this 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 Figure 2.

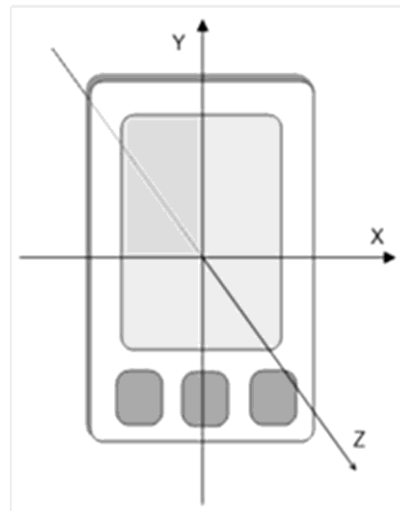


Figure 2: Axes of acceleration relative to the phone

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. Samples of raw tri-axial accelerometer data for simple and complex activities can be found in Pages 7 and 8 respectively. The orientation data is left as is. From the raw accelerometer data for simple activities, it could be clearly seen that every simple activity has a distinct pattern of acceleration on three different axes. However, intuitively such distinctive pattern is difficult to be extracted from the raw data of the complex activities. This poses a

difficulty in complex activity recognition which would be explained in Section IV.

Standard classifiers do not work well on the raw sensor data. It is essential to transform the raw data into a representation that captures the salient characteristics of the raw data. This is typically performed by breaking the continuous data into windows of certain duration. In this work we experimented with one, two, four, eight, twelve and sixteen seconds time windows. 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 [4]. We then extracted a number of features (as listed in Table I) to encode each window. Thus each window was represented as a feature vector of length 30.

These features were then used to train the classifiers.

D. Classification

The WEKA machine learning toolkit [6] was used to test classifiers using the features extracted from the raw data set. Six different classifiers were tested: Multi-layer Perceptron, Naïve Bayes, Bayesian network, Decision Table, Best-First Tree, and K-star. The accuracy of the classifiers was tested using ten-fold cross-validation with. Data collected from all the subjects was pooled together. The train and test data for each fold was randomly drawn from this pool, ensuring no overlap between the train and test sets of each fold. We used this approach as it is a good estimator for the generalized performance of the different classifiers. We used the default parameters associated with each of the classifier.

TABLE I. DESCRIPTION OF THE FEATURES EXTRACTED FROM THE RAW DATA

Feature	Accelerometer	Orientation	Description
Mean	X, Y, Z	Azimuth, Pitch, Roll	Average acceleration values along the three axes and average orientation along the three directions
Min	X, Y, Z	Azimuth, Pitch, Roll	The minimum acceleration and orientation value within the window along three axis of acceleration and directions of orientation
Max	X, Y, Z	Azimuth, Pitch, Roll	The maximum acceleration and orientation value within the window along the three axis of acceleration and directions of orientation
Standard Deviation	X, Y, Z	Azimuth, Pitch, Roll	The standard deviation in the acceleration and orientation values within the window
Zero-Cross	X, Y, Z		The number of zero crossings for the three axis of acceleration.
Correlation	X/Y, X/Z, Y/Z		The pairwise correlation between the three axes of acceleration.

TABLE II. CONFUSION MATRIX FOR COMBINED ACTIVITY SET (MULTI-LAYER PERCEPTRON, 1 SECOND WINDOW)

Classified as	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
Biking = a	1056	52	29	12	20	20	9	5	21	30	66	26	169	22	50
Climbing = b	28	694	18	18	12	33	21	33	31	11	7	12	28	8	13
Driving = c	3	25	3228	2	3	5	1	42	11	0	12	1	2	1	5
Lying = d	27	34	10	1413	9	2	2	1	5	2	4	16	2	2	1
Not-on-person = e	0	3	1	1	3360	0	1	1	0	0	0	0	1	0	0
Running = f	1	16	3	4	1	780	0	1	1	4	5	3	16	7	5
Sitting = g	27	51	29	6	1	6	1107	5	7	8	1	4	4	10	2
Standing = h	1	6	0	5	3	6	0	748	0	0	1	1	1	0	2
Walking = i	9	68	10	7	2	8	17	1	1472	8	13	14	40	2	22
Cleaning Kitchen = j	0	0	1	0	0	1	0	0	8	190	243	108	209	2	11
Cooking = k	1	1	0	5	0	4	1	0	5	123	774	160	448	18	123
Medication = l	0	5	0	1	0	3	0	1	19	77	125	1103	381	27	53
Sweeping = m	0	2	0	2	0	6	0	0	11	116	384	176	1257	21	68
Washing hands = n	0	1	1	0	1	2	0	0	4	42	96	62	149	56	38
Watering plants = o	0	0	0	1	1	8	2	0	8	30	125	70	171	24	245

TABLE III. DETAILED ACCURACY BY CLASS FOR COMPLEX ACTIVITIES (MULTILAYER PERCEPTRON, 1-SECOND WINDOW)

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

IV. RESULTS

We categorized the activities into three groups: simple, complex and combined to study the ability of the different classifiers to recognize them separately. The classification accuracies obtained from the different classifiers are summarized in Figure 3. It can be observed that the classification accuracies for simple activities remain consistently above 90% except for Naïve Bayes. However adding complex activities to the set reduced the performance of all the classifiers uniformly. The best accuracy noted for complex activities was 50% (guessing the class by chance stands at 17%). The best accuracy was obtained with the Multi-Layer Perceptron.

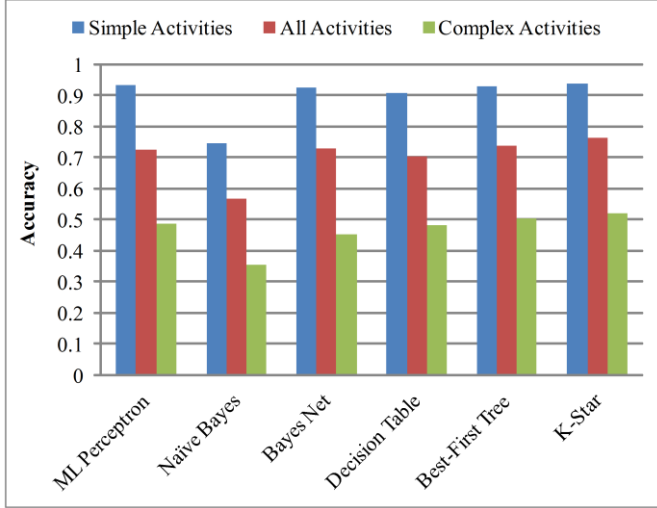


Figure 3: Performance of Different Classifiers

The confusion matrix for the combined set of activities is presented in Table II. The matrix shows how heavily the misclassification leans towards complex activities. This is not surprising alone but does show the resilience of simple activities against being misclassified when more activities are added. To gain further insights into the performance on these complex activities, we computed additional measures such as precision, recall and F score that is summarized in Table III. 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 also involved little movement in the experiment while the latter (with the exception of washing hands) contained much more movement from one area to another.

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.

Our next experiment studied the effect of different window lengths on the ability of the classifiers to recognize the activities. We used K-star as the classifier as it resulted in the

best performance in the previous experiment. The results of this experiment are summarized Figure 4. It can be noted that the overall classification accuracy for simple activities remains above 90% for each of the different window lengths. 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 activity.

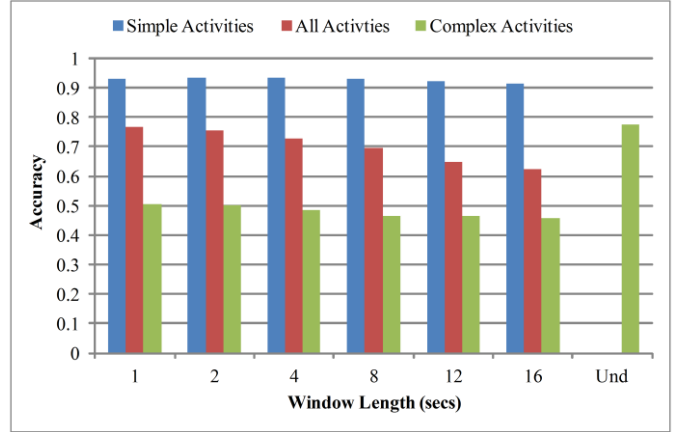


Figure 1: Classification accuracies for K-Star with different window lengths. Und corresponds to the scenario where train models for recognizing complex activities without using the sliding window protocol.

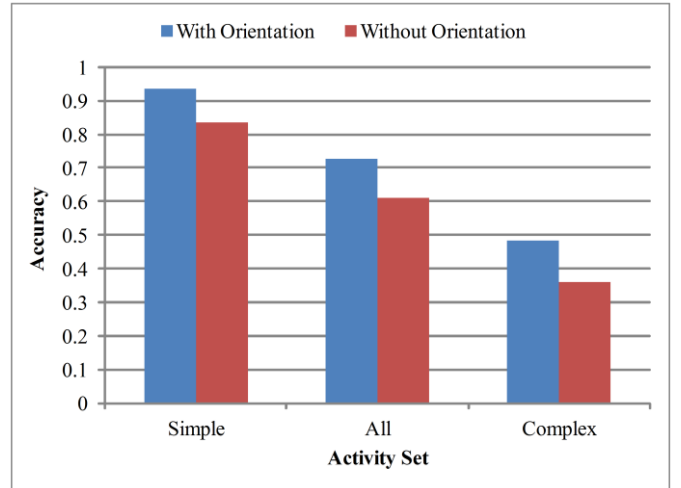


Figure 5: Accuracy of K-Star with and without using orientation information from gyroscope data

Figure 5 shows the results of classifying each activity set with and without orientation features extracted from gyroscope data. 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 study, orientation data represented an average of a 10-12% increase in accuracy over pure acceleration data. Given the format of 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.

This approach for recognizing activities does have some limitations. Clearly, the methodology is feasible for recognizing simple activities. However, it fares poorly in the context of complex activities. The process was made 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.

V. CONCLUSIONS

Simple activities can be recognized with very high accuracy using only a single smart phone carried naturally. Performance was over 93% using a Multi-layer Perceptron 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 at par with previous work on body sensors [3]. 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 [2]. 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.

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 phone's power usage. 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. 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 combination 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.

ACKNOWLEDGMENTS

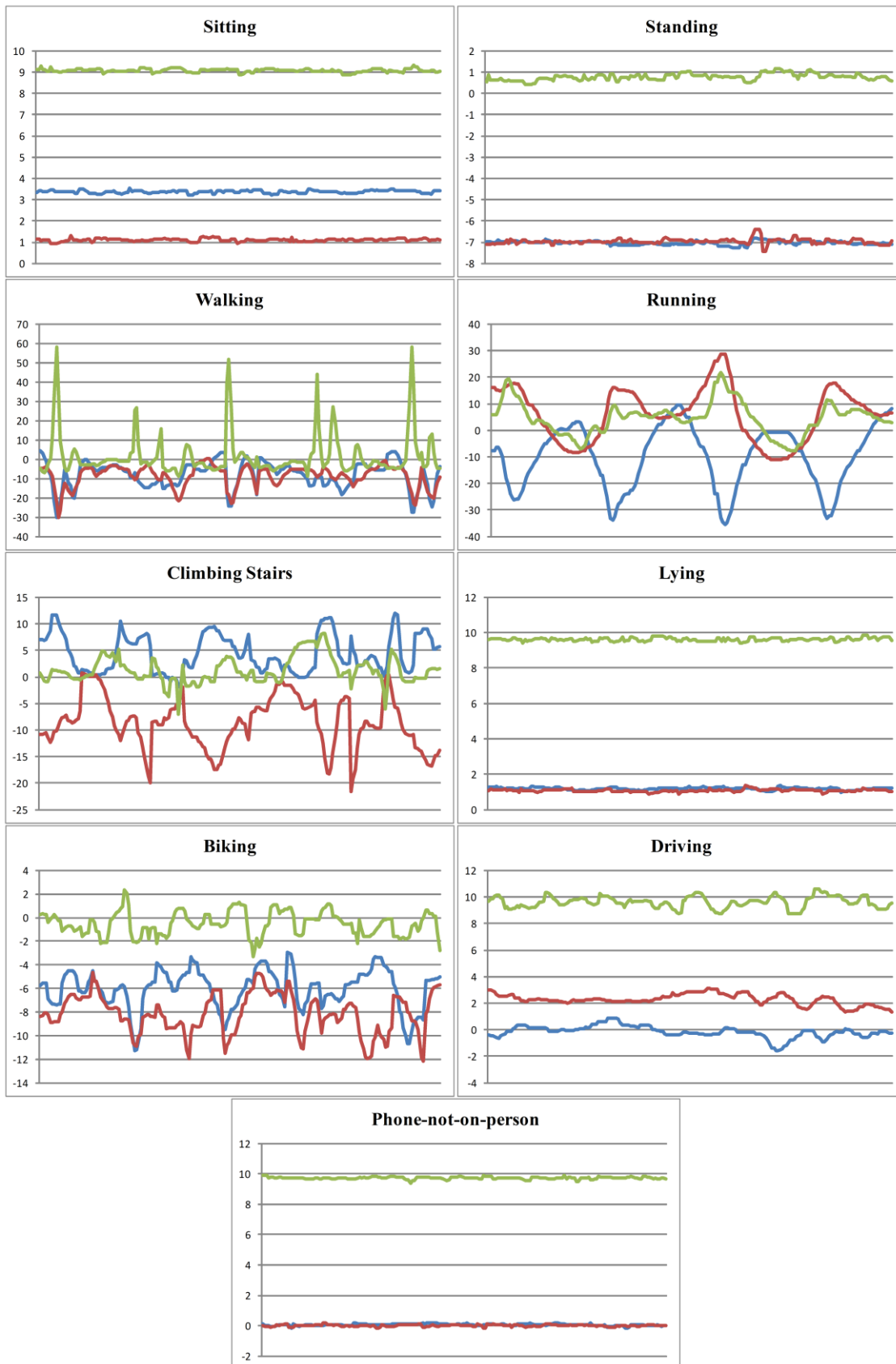
This work was supported by the National Science Foundation's REU program under grant number IIS-0647705.

REFERENCES

- [1] E. M. Tapia, S. S. Intille, and K. Larson, "Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors," *Proceedings of PERVASIVE*, pp. 158-175, 2004.
- [2] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz, and D. Hahnel, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, vol. 3, pp. 50-57, 2004.
- [3] D. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intelligent Systems*, 2012. (*To appear*)
- [4] L. Bao, S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," *Pervasive Computing, LNCS 3001*, pp. 1-17, 2004.
- [5] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity Recognition from Accelerometer Data," *Proceeding of the National Conference on Artificial Intelligence*, vol. 20, PART 3, pp. 1541-1546, 2005.
- [6] I. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," Morgan Kaufmann, 1999.
- [7] Tapia, E.M., S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-Time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor", In *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, pp.37-40, 2007.
- [8] N. C. Krishnan, D. Colbry, C. Juillard, and S. Panchanathan, "Real time human activity recognition using tri-Axial accelerometers", In *Sensors, Signals and Information Processing Workshop*, 2008.
- [9] N. Krishnan, and S. Panchanathan, "Analysis of Low Resolution Accelerometer Data for Continuous Human Activity Recognition", In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp3337-3340, 2008.
- [10] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions". In *IEEE Proceedings on the International Workshop on Wearable and Implantable Sensor Networks*, 2006.
- [11] S. W. Lee, and K. Mase, "Activity and location recognition using wearable sensors", In *IEEE Pervasive Computing*, 1(3):24-32, 2002.
- [12] A. Subramanya, A. Raj, J. Bilmes, and D. Fox, "Recognizing activities and spatial context using wearable sensors", In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 2006.
- [13] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers". In *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*, Washington DC, pp10-18, 2010.
- [14] J. Yang, "Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones", In *First International Workshop on Interactive Multimedia for Consumer Electronics at ACM Multimedia*, 2009.
- [15] T. Brezmes, J. L. Gorricho, and J. Cotrina, "Activity Recognition from accelerometer data on mobile phones", In *IWANN Proceedings of the 10th International Conference on Artificial Neural Networks*, pp.796-799, 2009.
- [16] G. Hache, E. D. Lemaire, N. Baddour, "Mobility change-of-state detection using a smartphone-based approach," In *Proceedings of International Workshop on Medical Measurements and Applications*, pp.43-46, 2010.
- [17] A. M. Khan, Y. K. Lee, S. Y. Lee, T. S. Kim, "Human Activity Recognition via an Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis," *5th International Conference on Future Information Technology*, pp.1-6, 2010.
- [18] S. Zhang, P. J. McCullagh, C. D. Nugent, and H. Zheng, "Activity Monitoring Using a Smart Phone's Accelerometer with Hierarchical Classification". In *Proceedings of the 6th International Conference on Intelligent Environments*, pp. 158-163, 2010.

Simple Activities

— X-axis — Y-axis — Z-axis



Complex Activities

— X-axis — Y-axis — Z-axis

