Cell Phone-Based Biometric Identification

Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore

Abstract— **Mobile devices are becoming increasingly sophisticated and now incorporate many diverse and powerful sensors. The latest generation of smart phones is especially laden with sensors, including GPS sensors, vision sensors (cameras), audio sensors (microphones), light sensors, temperature sensors, direction sensors (compasses), and acceleration sensors. In this paper we describe and evaluate a system that uses phone-based acceleration sensors, called accelerometers, to identify and authenticate cell phone users. This form of behavioral biometric identification is possible because a person's movements form a unique signature and this is reflected in the accelerometer data that they generate. To implement our system we collected accelerometer data from thirty-six users as they performed normal daily activities such as walking, jogging, and climbing stairs, aggregated this time series data into examples, and then applied standard classification algorithms to the resulting data to generate predictive models. These models either predict the identity of the individual from the set of thirty-six users, a task we call user identification, or predict whether (or not) the user is a specific user, a task we call user authentication. This work is notable because it enables identification and authentication to occur unobtrusively, without the users taking any extra actions—all they need to do is carry their cell phones. There are many uses for this work. For example, in environments where sharing may take place, our work can be used to automatically customize a mobile device to a user. It can also be used to provide device security by enabling usage for only specific users and can provide an extra level of identity verification.**

I. INTRODUCTION

Mobile devices, such as cellular phones, music players, and portable computers have recently begun to incorporate diverse and powerful sensors. These sensors include GPS sensors, audio sensors (microphones), image sensors (cameras), light sensors, temperature sensors, direction sensors (compasses) and acceleration sensors. Because of the small size of these "smart" mobile devices, their substantial computing power, their ability to send and receive data, and their nearly ubiquitous use in our society, these devices open up exciting new areas for data mining research and applications. The goal of the WISDM (Wireless Sensor Data Mining) project [1], which the work described in this paper is part of, is to explore the research issues and applications related to mining sensor data from these powerful mobile devices. In this paper we explore the use of one of these sensors, the accelerometer, in order to accomplish the biometric tasks of person identification and authentication.

We chose Android-based cell phones as the platform for our WISDM project because the Android operating system is free, open-source, easy to program, and is expected to quickly become a dominant entry in the cell phone marketplace. Our project employs several types of Android phones, including the Nexus One, HTC Hero, and Motorola Backflip. All of these Android phones, as well as virtually all new "smart" phones and "smart" music players, including the iPhone and iPod Touch [2], contain tri-axial accelerometers that measure acceleration in three spatial dimensions. Accelerometers were initially included in these devices to support advanced game play and to enable automatic screen rotation but, as we show in the paper, can be used for other purposes.

In this work we use accelerometer data to identify or authenticate cell phone users. This data is generated while the users perform normal daily activities, such as walking and climbing stairs, while keeping a cell phone in their pocket. Clinical research has shown that gait, the way a person walks or runs, is sufficient to distinguish between individuals [3] and our research confirms this. While we measure user movements using an accelerometer, other studies have focused on video images of users, and this alternate method has also been shown to be effective for identification [4]. While some previous work has examined sensor-based gait recognition [5-10], our work differs in that we identify users based on the way they move during multiple activities (i.e., not just walking) using only commercially available smart phones, which are carried in the user's pocket.

Once we have sample data from each of the users, we aggregate this raw time-series accelerometer data into examples, since most classification algorithms cannot operate directly on time series data (although we do plan to try Markov Models in the future). Each of these examples is associated with a specific cell phone user, thus forming labeled training data. For *user identification* we use this training data to build a single predictive model to match each example with a particular user. For *user authentication* we build separate models for each user in order to determine whether an example came from that user or from someone else. Using such a model, we are able to predict whether or not a user is who he or she claims to be.

Cell phone-based biometrics offers a wide range of possible applications. It can be used to authenticate users and thus to provide device security and theft prevention. Since other mobile devices, like laptops and music players also have accelerometers, they can also use this mechanism to provide theft prevention. This capability can also be used to provide an extra layer of security when one's identity must be authenticated. Finally, this form of biometric identification can be used to automatically personalize mobile devices—for example, music players that

can automatically configure particular settings or select certain songs after identifying the current user of the device.

Our work makes several contributions. One contribution concerns the accelerometer data that we have collected and continue to collect, which we ultimately plan to make available to other researchers. We also demonstrate how raw time series accelerometer data can be transformed into examples, which can then be used by conventional classification algorithms. Most significantly, we show that it is possible to perform identification and authentication with commonly available (nearly ubiquitous) equipment and yet achieve highly accurate results.

The remainder of this paper is structured as follows. Section II describes data collection and the process for transforming the time-series data into examples. Section III describes our experiments and results. Related work is described in Section IV and Section V summarizes our conclusions and discusses areas for future research.

II. DATA COLLECTION AND TRANSFORMATION

In this section we describe how we collect the accelerometer data and transform it into examples that can be used to build predictive models for user identification and authentication. For *user identification* we build a single predictive model from this training data, which will map each example to one of the thirty-six participants in our study. For *user authentication* we build an authentication model for each user, *u*. In order to train the authentication model for *u*, we map the class labels to just two values—*u* or *not-u*—and then generate the predictive model. The remainder of this section describes the data collection process and the aggregation process for transforming the raw time-series accelerometer data into examples for these tasks.

A. Data Collection

In order to collect data for our supervised learning tasks, it was necessary to have a large number of users carry an Android-based smart phone while performing certain everyday activities. Before collecting this data, we obtained approval from the Fordham University IRB (Institutional Review Board) since the study involved "experimenting" on human subjects and there was some risk of harm (e.g., the subject could trip while jogging or climbing stairs). We then enlisted the help of thirty-six volunteer subjects to carry a smart phone while performing a specific set of activities. These subjects were asked to walk, jog, climb up stairs, and climb down stairs for specific periods of time while they carried the Android phone in their front pants leg pocket. Data collection was stopped when the subjects switched activities. Data was collected in this manner so that it could also be used for another research study of ours that addresses activity recognition—a task that requires us to predict the activity that a user is performing [14]. Although such a scenario is somewhat less realistic when examining our biometrics tasks, we believe it is reasonable to expect a user to perform any one of these activities (walking, jogging, climbing stairs, or descending stairs) continuously for short periods of time.

Data collection was controlled by an Android application that we created that runs on the phones. This application, through a simple graphical user interface, permitted us to record the user's name, start and stop the data collection, and label the activity being performed. Through our application we could control what data was collected (e.g., GPS as well as accelerometer data) as well as how frequently it was collected. In all cases we collected the accelerometer data using our default sampling frequency of 50ms, or 20 samples per second. The data collection was supervised by one of the WISDM team members to ensure the quality of the data. Data was then collected wither directly from the phone via a USB connection or transmitted over a cellular connection to our Internet-connected server.

B. Feature Extraction

We first transform the raw time series data into examples, since the classification algorithms that we use in this paper cannot directly learn from time series data [11]. To accomplish this we divided the data into 10-second segments and then generated features from the accelerometer values contained in each 10-second interval (since acceleration data is collected for 3 axes 20 times per second for a 10-second interval there are 600 total values). We refer to this 10-second interval as the example duration (ED). We chose a 10-second example duration because we felt that it provided sufficient time to capture several repetitions of the (repetitive) motions involved in the four activities that we consider. Although we have not performed experiments to determine the optimal example duration value, we did compare the results for a 10-second and 20-second ED and the 10-second ED yielded slightly better results.

Next we generated informative features based on the 600 raw accelerometer readings. We generated a total of forty-three features, although these are variations of just six basic features. The forty-three features are described below, with the number of features generated for each feature-type noted in brackets:

- Average[3]: Average acceleration value (for each axis)

- Standard Deviation[3]: Standard deviation (for each axis)

- Average Absolute Difference[3]: Average absolute difference between the value of each of the 200 readings within the ED and the mean value over those 200 values (for each axis)

- Average Resultant Acceleration[1]: Average of the square roots of the sum of the values of each axis squared $\sqrt{(x_i^2 + y_i^2 + z_i^2)}$ over the ED

- Time Between Peaks[3]: Time in milliseconds between peaks in the sinusoidal waves associated with most activities (for each axis)

- Binned Distribution[30]: We determine the range of values for each axis (maximum – minimum), divide this range into 10 equal sized bins, and then record the fraction of the 200 values that fall within each of the bins.

The "time between peaks" feature requires further explanation. Repetitive activities, like walking, tend to generate repeating waves for some or most of the axes, and this feature measures the time between successive peaks. To estimate this value we find the highest peak within the record for each dimension, set a threshold based on a percentage of this value, and then find other peaks that met or exceed this threshold; if no peaks meeting this criterion are found then the threshold is lowered until we find at least three peaks. We then measure the time between successive peaks and calculate the average. For samples where at least three peaks could not be found, the distance between peaks is marked as unknown. This method was able to find the time between peaks for activities that had a clear repetitive pattern, like walking and jogging. Certainly more sophisticated schemes will be tried in the future.

The number of examples generated for each activity differs for each user, due to either time or physical limitations associated with the users or due to minor variations in the data collection process. The data set is summarized in Section III.

III. EXPERIMENTS

In this section we describe our experiments and then present our results. Our results for person identification are presented first and this is followed by our results for the authentication task.

A. Description of Experiments

The preparation required for our experiments involves collecting the raw accelerometer data and then transforming this time-series data into examples. This process was described in Section II. The resulting data set has 4,866 examples from thirty-six users, where each example contains forty-three features. This data set, which is subsequently used for training and testing, is described in Table 1. Due to space limitations we provide detailed statistics for only some of the thirty-six users.

TABLE 1. NUMBER OF EXAMPLES PER USER AND ACTIVITY

ID	Walk	Jog	Up	Down	Total
1	74	15	13	25	127
2	48	15	30	20	113
3	62	58	25	23	168
4	65	57	25	22	169
5	65	54	25	25	169
6	62	54	16	19	151
7	61	55	13	11	140
8	57	54	12	13	136
9	31	59	27	23	140
10	62	52	20	12	146
.	.	.	.	.	.
30	35	31	28	19	113
31	64	55	17	16	152
32	34	32	0	0	66
33	64	0	0	0	64
34	59	59	0	0	118
35	55	46	19	12	132
36	87	81	23	16	207
Sum	2081	1625	632	528	4866
%	42.8	33.4	13.0	10.8	100

The last two rows in Table 1 show the number and percentage, respectively, of the total examples associated with each activity. Certain activities contain fewer examples than others, mainly because the users were not asked to perform strenuous activities, like jogging and stair climbing, for very long.

We used the data set described in Table 1 to generate six distinct data sets for the person identification task. The "aggregate" data set uses examples from all four activities but with the activity label removed, so that the type of activity is not explicitly encoded in the data. This data set thus represents our most realistic scenario, because in real-world situations users will perform multiple activities and we will not expect them to explicitly label the activity they are performing. We also created four data sets, each of which contains examples from only one activity (walking, jogging, ascending stairs, and descending stairs). Results based on these data sets may be less practical, but they provide insight into how useful each activity is for discriminating between users. Finally, our sixth data set, which we refer to as "aggregate (oracle)," is identical to the aggregate data set but contains the activity label as a feature. While this data set does not correspond to a very realistic scenario, it provides us with information on the utility of knowing the class label. This is important because, as our previous research on activity recognition has demonstrated [14], we can accurately identify most activities using the same methods used in this paper to identify the users. Thus, should the activity labels provide a substantial benefit, our aggregate (oracle) data set will provide an upper bound on the performance of a two-stage learning approach, where we first predict the activity and then use this to help identify the user.

After preparing these six data sets, we used two classification techniques from the WEKA data mining suite [12] to induce models for person identification—decision trees (J48) and neural networks. In each case we used the default settings. Thus, twelve experiments in identification were performed. We used ten-fold cross validation for all experiments and all results are based on these ten runs.

We addressed the authentication task by transforming the multi-class identification problem into a binary classification problem, where the positive class corresponds to the user to be authenticated and the negative class to the other (thirty-five) users. Because the positive class is so rare (on average it contains 1/36 of the data), most classification methods will tend to generate classifiers that do not perform well at predicting this (rare) class [15]. Since this is not desirable in this situation, we under-sampled the negative class such that the resulting ratio of positive examples to negative examples is 1:3 (i.e., the person to be authenticated makes up 25% of the examples). We did not use a 1:1 ratio, which is often used when learning from unbalanced data [15], because of the limited amount of data and our desire to have a reasonably sized training set.

The authentication data set used examples from all activities (i.e., we used the aggregate data set). A separate authentication model is required for each user, but due to time constraints we generated authentication models for only the first five users in our data set. As with person identification we used ten-fold

cross validation to build and evaluate each authentication model. We used J48 to build these models. Our authentication results are described per class so that the fact that we changed the class distribution does not impact or distort the results.

B. Identification Results

In this section we present our results for the identification task. Table 2 shows the results for the six data sets introduced earlier in this section when using WEKA's J48 and Neural Network learning algorithms, as well as when using our "Straw Man" strategy. The straw man strategy is for comparison purposes and corresponds to the strategy of always predicting the most common class (i.e., user). Finally, recall that these results are based on individual test examples, which are formed from just ten seconds worth of data. As we shall soon see, if we base identification on multiple samples, then it is possible to improve the identification performance.

TABLE 2. ACCURACIES (%) FOR IDENTIFICATION OF 10-SECOND EXAMPLES

	Aggregate	Walk	Jog	Up	Down	Aggregate (Oracle)
J48	72.2	84.0	83.0	65.8	61.0	76.1
Neural Net	69.5	90.9	92.2	63.3	54.5	78.6
Straw Man	4.3	4.2	5.0	6.5	4.7	4.3

The results in Table 2 indicate that our models are quite successful at recognizing user's identities based upon only 10 seconds of accelerometer data. While some of the accuracies may not seem that good, they are actually quite impressive when one considers that for this 36-class classification problem the straw man strategy of always guessing the most frequent class yields an accuracy in the 4-7% range. Using J48, the aggregate data set achieved 72.2% accuracy, which indicates that such models could be plausible in realistic scenarios. Note that for both J48 and the Neural Network the aggregate (oracle) data set achieved only modestly higher accuracy than the aggregate data set, indicating that it is not critical to know what activity a user is performing in order to identify a user. The walking and jogging data sets also show the best overall performance and perform better than the aggregate data set, indicating that, if we do have control over the users, walking is a good activity to use for identification purposes (jogging is probably not practical given that not everyone is fit enough to jog). Ascending and descending stairs have substantially lower accuracies, possibly because these activities have much less data available for building the model (see Table 1).

We report the identification accuracies per user in Table 3, again based on the performance for each 10-second example. In the interest of space we present the results only for the aggregate, walking, and jogging activities. In some cases there were no examples for an activity and this is denoted with a "–" in the cell. As Table 3 indicates, the accuracies are generally quite high when considering that the straw man strategy of guessing the most frequent user yields accuracies in the range of 4-7%.

TABLE 3. ACCURACIES (%) FOR IDENTIFICATION OF 10-SECOND EXAMPLES

	Aggregate		Walking		Jogging	
	J48	NN	J48	NN	J48	NN
1	66.9	65.4	87.8	98.7	40	40.0
2	72.6	77.9	77.1	100.0	60	93.3
3	69.1	81.0	74.2	96.8	82.8	96.6
4	75.7	69.2	93.9	96.9	87.7	96.5
5	74.0	69.8	87.7	95.4	90.7	96.3
6	78.2	70.9	90.3	96.8	83.3	100.0
7	61.4	61.4	73.78	82.0	72.7	63.6
8	75.7	77.2	86.0	98.3	74.1	85.2
9	72.6	78.1	83.9	98.4	86.5	94.2
10	70.0	45.7	77.4	96.8	86.4	98.3
11	72.2	84.0	87.5	96.9	90.9	92.7
12	76.8	54.6	88.9	83.3	81.0	84.1
13	70.2	59.0	86.7	51.7	79.0	98.4
14	66.2	77.9	87.1	96.8	–	–
15	62.8	71.3	91.8	85.3	81.3	96.9
16	80.0	61.2	89.2	98.5	95.1	96.7
17	82.0	87.0	87.1	98.6	–	–
18	84.2	88.5	81.8	97.0	88.1	96.6
19	75.9	71.2	88.4	95.7	75.8	97.0
20	56.5	47.6	77.4	87.1	77.4	96.9
21	70.6	54.1	85.3	94.4	95.2	96.8
22	75.2	72.9	87.9	97.0	96.7	96.7
23	43.4	54.7	66.7	93.3	0	60
24	85.9	93.4	95.2	98.4	–	–
25	77.4	75.6	91.0	97.0	92.2	96.9
26	75.2	72.3	78.8	92.9	86.5	92.3
27	67.8	53.3	77.4	95.2	78.6	95.7
28	71.7	83.8	96.9	100.0	94.7	100.0
29	67.5	63.7	70.8	81.5	92.7	61.8
30	87.6	49.6	88.6	97.1	90.3	96.8
31	71.1	77.0	82.8	93.8	65.5	90.9
32	71.2	83.3	91.3	85.3	62.5	96.9
33	67.2	45.3	78.1	53.1	–	–
34	77.1	89.8	84.8	88.1	89.8	96.6
35	51.5	45.5	74.6	87.3	52.2	84.8
36	73.9	82.1	79.3	74.7	92.7	96.3

We can achieve *vastly* superior results if we base identification on more than ten seconds worth of data. The strategy that we recommend involves taking multiple examples generated from a single user and then assigning the identity based on the user that is predicted most often. We refer to this as the *most frequent user* strategy. In our implementation of this strategy we consider *all* of the examples from a single user in the test set, but this typically corresponds to only about 5-10 minutes worth of data, which seems to be a reasonable length sample. Table 4 reports the fraction of the users that are correctly identified when using the most frequent user strategy. In some cases the denominator is less than 36 because we did not obtain data for all activities from all users. Fractional values are the result of averaging the results over 10 runs.

TABLE 4. USERS CORRECTLY IDENTIFIED USING MOST FREQUENT USER

	Aggregate	Walk	Jog	Up	Down	Aggregate (Oracle)
J48	36/36	36/36	31/32	31/31	28/31	36/36
Neural Net	36/36	36/36	32/32	28.5/31	25/31	36/36

Table 4 demonstrates that when using the most frequent user strategy we are able to achieve perfect identification accuracy in the most realistic situation—when using the aggregate data set. Even in cases where the accuracies for the 10-second examples are far from perfect, we can still accurately identify the users because, as our detailed analysis of the errors shows, the errors are distributed across many users and not concentrated on just one or two users—thus the most frequent user strategy had no problem correctly identifying the correct user. To demonstrate this we provide more detailed information about the errors in Table 5. Specifically, for each user we determined the user most often incorrectly predicted (i.e., the "imposter" most often confused with the actual user). We compared the number of these erroneous predictions with the number of times the user was correctly identified. This ratio of these values (correct to incorrect) for the aggregate data set is presented in Table 5.

TABLE 5.   RATIO OF CORRECTLY IDENTIFIED RECORDS TO INCORRECTLY IDENTIFIED RECORDS OF MOST FREQUENT IMPOSTER

User	J48	Neural Net	User	J48	Neural Net
1	14:1	12:1	19	24:1	9:1
2	21:1	15:1	20	10:1	2:1
3	15:1	27:1	21	34:1	5:1
4	21:1	4:1	22	24:1	23:1
5	25:1	15:1	23	4:1	3:1
6	30:1	21:1	24	22:1	49:1
7	11:1	10:1	25	26:1	18:1
8	15:1	18:1	26	20:1	7:1
9	27:1	23:1	27	16:1	9:1
10	8:1	5:1	28	17:1	16:1
11	17:1	13:1	29	17:1	7:1
12	25:1	4:1	30	24:1	2:1
13	16:1	16:1	31	18:1	16:1
14	5:1	20:1	32	6:1	18:1
15	16:1	46:1	33	14:1	5:1
16	17:1	13:1	34	30:1	26:1
17	21:1	17:1	35	6:1	7:1
18	17:1	49:1	36	13:1	12:1

The results in Table 5 show that in no case is an imposter almost mistaken for the true user (i.e., the ratios are significantly greater than 1:1). In the cases where the ratio is smallest, this is due to a limited amount of training data from the user; however, as seen in the case of the aggregate data set, even a user with very few examples like User 23 still has significantly more examples that are correctly identified than are incorrectly identified as the most frequent imposter.

The results in Table 5 are important because they suggest that our performance will scale up to significantly more than 36 users and that good identification performance is possible even if the accuracies associated with individual examples is only modest. In order to get more insight into the impact of the number of users on performance, we analyzed how the accuracy of individual predictions for the aggregate data set is affected by the total number of users. The results, which are shown in Figure 1, show that while accuracy does decrease with the number of users, the decrease is gradual once we have more than a few users. Our expectation is that with the most frequent user strategy we probably could maintain perfect identification accuracy even if we had two or three times as many users.
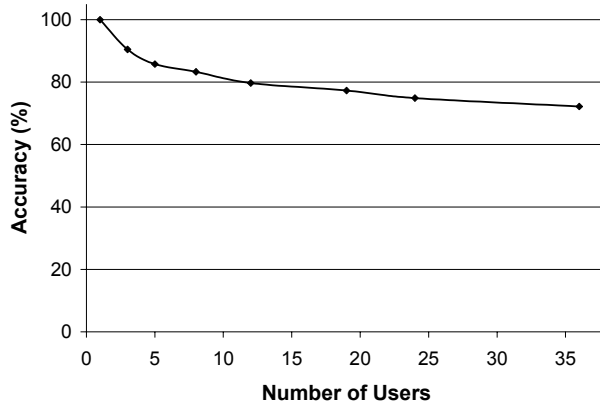

Fig. 1.   The identification accuracy associated with individual ten-second examples as the number of users included in the aggregate data set increases.

C.   Authentication Results

Our experiments in authentication obtained promising results as well. Because authentication requires a separate model for each user, we report authentication results for only five users due to time constraints. As with the case for person identification, we first present the results associated with individual examples and then apply a most frequent user strategy to determine the actual authentication performance statistics.

Our key statistics for authentication are the positive authentication rate and the negative authentication rate. To make things more concrete, here we assume that we are trying to authenticate user 1, $u1$. The positive authentication rate for $u1$ is the fraction of test examples coming from $u1$ that are correctly classified as belonging to $u1$. The negative authentication rate for $u1$, or the imposter identification rate, is the fraction of test examples from an imposter (i.e., not $u1$) that are correctly identified as not belonging to $u1$. Clearly we would like to have a high positive authentication rate and a high negative authentication/imposter identification rate.

Table 6 summarizes the authentication results. In addition, to showing the positive and negative authentication rates, it also shows the results of applying the most frequent user strategy to all of the individually classified test examples. For the most frequent user strategy we use a '+' to indicate that a user is correctly identified as himself or herself (for positive authentication) or that an imposter is correctly identified as an imposter (for negative authentication).

TABLE 6.   ACCURACIES (%) FOR IDENTIFICATION OF 10-SECOND EXAMPLES

User	Positive Authentication Rate (%)		Negative Authentication Rate (%)	
	10 sec. Example	Most Frequent User	10 sec. Example	Most Frequent User
1	86.6	+	94.8	+
2	92.9	+	96.5	+
3	82.1	+	94.2	+
4	85.8	+	97.1	+
5	82.2	+	92.6	+

Our results for these five users demonstrate that accelerometer data can be used to authenticate user's identities with a basic level of accuracy with only 10 seconds worth of data. However, these rates may not be sufficient by themselves for acceptable levels of real-world authentication—where we would need authentication rates well in excess of 99%. To achieve such rates we need to obtain multiple examples from each user and then apply the most frequent user strategy. In our experiments we achieve 100% positive and negative authentication rates for all five users. We believe that by using multiple test samples we can consistently achieve perfect or near perfect authentication rates.

IV. RELATED WORK

The use of sensor data for biometric identification and authentication is relatively new but has been increasingly explored in recent years. Gait recognition—the use of a person's unique style of walking to identify or authenticate one's identity—has shown some promising results as a biometrics tool. In his survey of biometric gait recognition, Gafurov [5] identifies three areas of gait recognition research—machine vision-based, floor sensor-based and wearable sensor-based methods. Our research focuses on the wearable sensor-based approach, which has been much less widely explored than the machine vision-based approach that uses camera images of users to identify them by their gaits [3]. While vision-based approaches have typically focused on gait while walking, Yam et al. [4] explored how gait during running *and* walking could be used to identify users. They found that gait while running was more useful than gait while walking for identification purposes, demonstrating the importance of utilizing multiple activities.

Previous work using the wearable sensor-based approach includes work by Mantyjarvi et al. [6], where users were identified via a small accelerometer-based device placed on a belt at the middle of the user's waistlines in back. Data was collected from thirty-six subjects who walked at fast, normal, and slow walking speeds. Correlation, frequency domain, and data distribution statistics were used for identification. Annadhorai et al. [7] performed experiments similar to the Mantyjarvi work. Their system used two wireless sensor nodes: a unit consisting of a tri-axial accelerometer and a bi-axial gyroscope and a base station unit which received sensor readings from other node. Data was collected from four users who wore the device above their ankles. Gafurov, Helkala, & Sondrol [8] authenticated user's identities using data collected from 21 users, using a tri-axial accelerometer-based device attached to the user's right lower legs. The readings for each axis were transformed into a combined gait signal, similar to our resultant acceleration feature. A similar procedure was followed by the same authors in another work [9], except that the mobile device used for data collection was a Micro-Electro-Mechanical-System (MEMS).

Gafurov & Snekkenes [10] analyzed data collected from accelerometer-based sensors placed on the foot, hip, pocket, and arm to authenticate user's identities. The accelerometer placed on the foot gave the best results. They also found that attributes such as weight of a shoe, sideways direction of ankle/foot mo-

tion, and different points of the gait cycle could impact recognition. Gait recognition was shown to perform well when an imposter attempted to mimic another person's gait; however, when the imposter knew the gender of the users in the data set or knew his closest match in the data set, the imposter's attempts were found to have higher chances of being wrongfully accepted. Gafurov, Snekkenes, and Buvarp [13] further supported these conclusions regarding the robustness again impersonation attacks in another work using an accelerometer-based MEMS placed on the hip in experiments with 20 and 22 users.

While these works offer important insight in the area of sensor-based gait recognition, our work differs in several important ways. Firstly, the devices used in our experiments, smart phones, are not specialized sensors but are widely available commercial devices that are routinely carried by millions of users. Furthermore, we relied only on the device being carried in the user's pocket—a natural location to carry such a device—while the other work typically involved the user being monitored with multiple sensors, often placed in awkward body locations. We also consider a set of users that is significantly larger than in some of the prior studies.

Another basic way that our work differs from previous work is that most sensor-based biometric exploration has focused only on analyzing gait while walking. However, the effect of multiple activities rather than simply walking is of interest because 1) different activities may provide a more robust biometric signature than just walking, 2) users perform many different activities during a day that a continuous identification or authentication systems may fail to recognize if the system is based only on walking gait recognition, and 3) it will be much easier to automatically collect movement data for training a model if the user is permitted to perform his daily routine. Yam et al. [4] used machine vision-based approaches to examine both running and walking movement for biometric identification, but our work goes further in that we analyzed four activities. We hope to expand the number of activities in the future and to collect data as users perform their daily routines.

V. CONCLUSIONS & FUTURE WORK

In this paper we described how a smart phone can be used to perform person identification and authentication, simply by keeping it in one's pocket. We have demonstrated that acceleration data collected while walking, jogging, ascending stairs, and descending stairs all have the potential to function as biometric signatures. In addition, we show that users can often be recognized quickly, using only 10 seconds worth of data. But we also showed that we can build highly accurate models for identification and authentication—and can often achieve perfect performance—if we utilize multiple samples before making an identification or authentication decision. We have also demonstrated that such models have the potential to perform well in realistic settings in which the activity that a user is performing is unknown and, unlike most prior work, our system requires no specialized equipment or awkwardly placed sensors.

We plan to improve our identification and authentication systems in several ways. With respect to data collection, we

intend to increase the number of users in the data set, collect more data per user, and expand the number of activities for which we collect data. In addition and perhaps most significantly, in the future we plan to collect data from the users in a more natural manner. Instead of having a researcher run the user through a well defined course with specific activities, all of the accelerometer data from the user's cell phone will be collected, without knowledge of what activity is being performed. Thus we will learn to identify users using profiles that can be generated in a completely automated manner, without the users taking any specific actions. Our aggregate data set currently approximates this scenario, but the approximation is far from perfect (e.g., only four activities). We will also try methods for classification that can operate directly on the time-series data, like hidden Markov models. Such methods can potentially yield superior results since information is lost when the raw time-series data is aggregated into examples.

Our work would not have been possible without establishing our WISDM Android-based data collection platform, and we view this software and hardware architecture, in which data is transmitted by the phone to our Internet-based server, as a key resource for future work. This platform, as well as the data that is collected, will ultimately be made public. We also plan to significantly enhance our WISDM platform so that we can generate results in real-time, whereas currently our results are generated off-line and are not reported back to the mobile phone and user. We plan to provide real-time results using two approaches. The first approach minimizes the intelligence required on the phone by having the phone transmit the data to our Internet-based sever over the cellular connection, with the server applying the biometrics model and transmitting the results back to the phone. The second approach involves implementing the model directly on the phone. This is feasible given the computational power of these devices and has the advantage of not requiring a server or any shared resources, which makes the system infinitely scalable, and also ensures the user's privacy, since the sensor data then can be kept locally on the user's cell phone.

The work described in this paper is part of a larger effort to mine sensor data from wireless devices. We plan to continue our WISDM project and will apply the accelerometer data to other tasks. In addition, we plan to collect and then mine other sensor data, including GPS data. We believe that mobile sensor data provides tremendous opportunities for data mining and we intend to leverage our Android-based data collection/data mining platform to the fullest extent possible.

REFERENCES

[1] WISDM Project. Fordham University, Department of Computer and Information Science, http://storm.cis.fordham.edu/~gweiss/ wisdm/
[2] Apple iPhone & Apple iPod Touch. Apple Inc., 2009. www.apple.com.
[3] M. Nixon, T. Tan, and R. Chellapppa, *Human Identification Based on Gait.* New York: Springer Science + Business Media Inc., 2006, ch. 1.
[4] C.Y. Yam, M. Nixon, and J. Carter, "Automated Person Recognition by Walking and Running via Model-based Approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1-16, 2003.
[5] D. Gafurov, "A Survey of Biometric Gait Recognition: Approaches, Security, & Challenges," *Annual Norwegian Computer Science Conference*, Oslo, Norway, November 19-21, 2007.
[6] J. Mantyjarvi, M. Lindholdm, E. Vildjounaite, S.M. Makela, and H. Ailisto, "Identifying Users of Portable Devices from Gait Pattern with Accelerometers," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp. 973-976, 2005.
[7] A. Annadhorai, E. Gutenberg, J. Barnes, K. Harage, and R. Jafari, "Human Identification by Gait Analysis," *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, Breckenridge, Colorado, June 17, 2008.
[8] D. Gafurov, K. Helkala, and T. Sondrol, "Biometric Gait Authentication Using Accelerometer Sensor.," *Journal of Computers*, vol. 1, no. 7, pp. 51-59, 2006.
[9] D. Gafurov, K. Helkala, and T. Sondrol, "Gait Recognition Using Acceleration from MEMS," *Proceedings of the First International Conference on Availability, Reliability, and Security,* pp. 6-12, 2006.
[10] D. Gafurov and E. Snekkenes, "Gait Recognition Using Wearable Motion Recording Sensors," *EURASIP Journal on Advances in Signal Processing*, 2008.
[11] G. M. Weiss and H. Hirsh, "Learning to predict rare events in event sequences" *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, CA, pp. 359-363, 1998.
[12] Witten, I. and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco, Morgan Kaufmann Publishers, 2005.
[13] D. Gafurov, E. Snekkenes and T.E. Buvarp, "Robustness of Biometric Gait Authentication Against Impersonation Attack," *International Workshop on Information Security, OnTheMove Federated Conferences*, Springer LNCS 4277, Montpellier, France, Oct 29 - Nov 3, 2006.
[14] J.R. Kwapisz, G. M. Weiss, and S.A. Moore. "Activity recognition using cell phone accelerometers," Proceedings of the *Fourth International Workshop on Knowledge Discovery from Sensor Data,* pp. 10-18, 2010.
[15] G. M. Weiss and Foster Provost, "Learning when training data are costly: The effect of class distribution on tree induction, *Journal of Artificial Intelligence Research*, vol. 19, pp. 315-354, 2003.